Coarse-graining Kohn-Sham Density Functional Theory

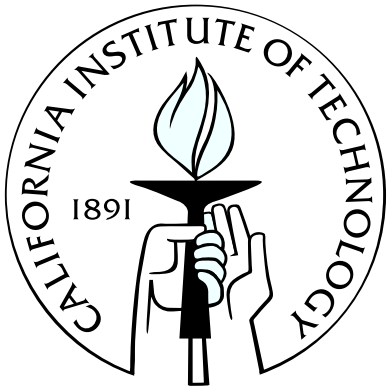
Thesis by

Phanish Suryanarayana

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2011

(Defended May 24, 2011)

Dedicated to my parents and Deepa

Acknowledgements

I would like to take this opportunity to thank all the people who have made my stay at Caltech memorable and pleasurable. First and foremost, I would like to thank my advisors Prof. Kaushik Bhattacharya and Prof. Michael Ortiz who provided me with an interesting and challenging problem for my thesis. Further, their guidance in the research methodology was invaluable. I would also like to thank Prof. Guruswami Ravichandran who provided me excellent advice through my Masters and Ph.D. I am also grateful to Prof. Nadia Lapusta and Prof. Chiara Daraio for serving on my committee and providing feedback on my thesis.

I have been fortunate to have been part of two research groups. I would like to acknowledge all my group members. In particular, I would like to thank Alex Kelly, Vikram Gavini, Thomas Blesgen, Luigi Perotti, Celia Reina Romo, Bharat Prasad, Andrew Richards, Malena Espanol and Agustin Bompadre. I would also like to thank all my friends, who have made my time at Caltech unforgettable. In particular, I would like to thank Naresh Satyan, who has been my roommate for 5 years. We have had some amazing experiences together. I would also like to thank Tejaswi Navilarekallu, Chinthaka Mallikarachchi, Srivatsan Hulikal, Ahmed Elbanna, Shankar Kalyanaraman, Shaunak Sen and Vikram Deshpande.

I am grateful to have a wonderful family. I would like to thank my parents and sister, who have

always supported me in my endeavors. Last but not the least, I would like to acknowledge Deepa Nagendra, with whom I have shared some memorable and remarkable experiences.

Abstract

Defects, though present in relatively minute concentrations, play a significant role in determining macroscopic properties. Even vacancies, the simplest and most common type of defect, are fundamental to phenomena like creep, spall and radiation ageing. This necessitates an accurate characterization of defects at physically relevant concentrations, which is typically in parts per million. This represents a unique challenge since both the electronic structure of the defect core as well as the long range elastic field need to be resolved simultaneously. Unfortunately, accurate ab-initio electronic structure calculations are limited to a few hundred atoms, which is orders of magnitude smaller than that necessary for a complete description. Thus, defects represent a truly challenging multiscale problem.

Density functional theory developed by Hohenberg, Kohn and Sham (DFT) is a widely accepted, reliable ab-initio method for computing a wide range of material properties. We present a real-space, non-periodic, finite-element and max-ent formulation for DFT. We transform the original variational problem into a local saddle-point problem, and show its well-posedness by proving the existence of minimizers. Further, we prove the convergence of finite-element approximations including numerical quadratures. Based on domain decomposition, we develop parallel finite-element and max-ent implementations of this formulation capable of performing both all-electron

and pseudopotential calculations. We assess the accuracy of the formulation through selected test cases and demonstrate good agreement with the literature.

Traditional implementations of DFT solve for the wavefunctions, a procedure which has cubic-scaling with respect to the number of atoms. This places serious limitations on the size of the system which can be studied. Further, they are not amenable to coarse-graining since the wavefunctions need to be orthonormal, a global constraint. To overcome this, we develop a linear-scaling method for DFT where the key idea is to directly evaluate the electron density without solving for the individual wavefunctions. Based on this linear-scaling method, we develop a numerical scheme to coarse-grain DFT derived solely based on approximation theory, without the introduction of any new equations and resultant spurious physics. This allows us to study defects at a fraction of the original computational cost, without any significant loss of accuracy. We demonstrate the efficiency and efficacy of the proposed methods through examples. This work enables the study of defects like vacancies, dislocations, interfaces and crack tips using DFT to be computationally viable.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Schrödinger equation is fundamental for describing the quantum mechanical electronic structure of matter, since it does not require any empirical input. However, the solution of the Schrödinger equation is exceedingly expensive, and this limits the size of systems that can be directly evaluated to tens of electrons (Kohn (1999)). Numerous approaches have been proposed to reduce the computational cost of the solution of the Schrödinger equation. These approaches include the widely used DFT of Hohenberg and Kohn (1964). In their seminal work, Hohenberg and Kohn (1964) proved the existence of a one-to-one correspondence between the ground state electron density and the ground state wavefunction of a many-particle system. By this correspondence, the electron density replaces the many-body electronic wavefunction as the fundamental unknown field, thereby greatly reducing the dimensionality and computational complexity of the problem.

The most common present-day implementation of DFT is through the Kohn-Sham method (Kohn and Sham (1965)), in which the intractable many-body problem of interacting electrons is reduced to a tractable problem of non-interacting electrons moving in an effective potential. Hidden in this formulation is the unknown exchange and correlation functional for which various models including the local spin-density approximation (LSDA) (Kohn and Sham (1965)) and the

generalized gradient approximation (GGA) (Langreth and Mehl (1983); Perdew et al. (1992)) are used.

The plane-wave basis is one of the most-frequently used basis for solving the Kohn-Sham problem (Kresse and Furthmüller (1996); Segall et al. (2002); Gonze et al. (2002); Ismail-Beigi and Arias (2000)) because of a number of attractive features. It forms a complete and orthonormal set that is independent of the atomic positions and is efficient for evaluating convolutions through the fast Fourier transform (FFT). However, the plane-wave basis also suffers from a few notable disadvantages. Firstly, the plane-wave basis is best suited to periodic systems. Therefore, the study of non-periodic and localized systems such as defects, clusters and surfaces requires the introduction of artificial supercell periodicity, which can lead to spurious results. Secondly, the plane-wave basis functions are non-local in real space, resulting in dense matrices which are ill-suited to iterative solution schemes. In addition, non-locality limits the usefulness of a plane-wave basis in multiscale approaches formulated in real space. Over the past decade, numerous efforts have been directed towards the development of real-space DFT implementations that overcome these difficulties (e.g., Hehre et al. (1969); Wills and Cooper (1987); Soler et al. (2002); Skylaris et al. (2005); Bowler et al. (2006); Chelikowsky et al. (1994); Castro et al. (2006)). Though some of these studies use the finite-element basis (Pask et al. (1999); Tsuchida (2004)), they do not utilize the unstructured nature of the the finite-element method and the convergence of the finite-element approximation with numerical quadratures has not been rigorously justified.

The computational complexity of DFT poses a significant hurdle in the solution of large systems of interest—irrespective of the basis set. Various approaches have been proposed to overcome this limitation. A few examples include the integration of DFT with molecular dynamics (Car and

Parrinello (1985)), linear and sub-linear scaling algorithms (García-Cervera et al. (2007); Garcia-Cervera et al. (2009); Goedecker (1999); Galli and Parrinello (1992); Mauri et al. (1993); Skylaris et al. (2005)) and adaptive-coordinate systems (Gygi and Galli (1995); Zumbach et al. (1996); Modine et al. (1997)) which improve the efficiency of the real space calculations by selectively having some regions of space with higher resolution than others.

Inspite of these advances, the study of crystal defects at realistic concentrations remains intractable. Defects present a unique challenge since both the electronic structure of the core as well as the long range elastic field need to be simultaneously resolved. Multiscale approaches which coarse-grain DFT or embed it into simpler, less accurate models like tight-binding (TB) or empirical potentials have also been developed recently (Choly et al. (2005); Lu et al. (2006); Govind et al. (1999); Bernstein et al. (2009)). Though these multiscale methods provide valuable insight, they suffer from a few notable drawbacks. In some cases, there is no seamless transition from DFT to TB or empirical potentials, while in others, uncontrolled approximations made by the use of linear response theory or Cauchy Born hypothesis render them unreliable. Also, there is no systematic convergence of the solution of these models to the full DFT solution.

In this work, we first present a non-periodic, real-space, finite-element and *max-ent* formulation of DFT. We show the well-posedness of this formulation for both the all-electron problem as well as the pseudopotential approximation. In particular, we prove the existence of solutions, and, in addition, the convergence of finite-element approximations, including numerical quadratures, using Γ-convergence methods. Further, we develop a parallel implementation of this formulation capable of performing both all-electron and pseudopotential calculations. Next, we present a formulation to seamlessly coarse-grain DFT solely based on approximation theory, without the

introduction of any new equations or physics. This allows us to study defects at a fraction of the original computational cost, without any significant loss of accuracy. Additionally, we can obtain systematic convergence to the fully resolved solution. In formulating the method, we first develop a linear-scaling method where we circumvent the calculation of the wavefunctions and directly evaluate the electron density. This is necessary since wavefunctions are global quantities which are not amenable to coarse-graining. Next, we introduce coarse-graining approximations, whereby we have high resolution in the vicinity of the defect and progressively coarsen as we move away from the defect. Therefore, we are able to perform the calculations both accurately and efficiently.

The remainder of the thesis is organized as follows. In Chapter 2, we develop a non-periodic, finite-element formulation of DFT. In Chapter 3, we present a mesh-free convex approximation scheme for DFT. We present the proposed linear-scaling method for DFT in Chapter 4. Subsequently, we present the proposed formulation for coarse-graining DFT in Chapter 5. Finally, we conclude in Chapter 6 with a short discussion and scope for future work.

Chapter 2

Non-periodic finite-element formulation of Kohn-Sham density functional theory

The plane-wave basis is one of the most-frequently used basis for solving the Kohn-Sham problem (Kresse and Furthmüller (1996); Segall et al. (2002); Gonze et al. (2002); Ismail-Beigi and Arias (2000)) because of a number of attractive features. It forms a complete and orthonormal set that is independent of the atomic positions and is efficient for evaluating convolutions through the fast Fourier transform (FFT). However, the plane-wave basis also suffers from a few notable disadvantages. Firstly, the plane-wave basis is best suited to periodic systems. Therefore, the study of non-periodic and localized systems such as defects, clusters and surfaces requires the introduction of artificial supercell periodicity, which can lead to spurious results. Secondly, the plane-wave basis functions are non-local in real space, resulting in dense matrices which are ill-suited to iterative solution schemes. In addition, non-locality limits the usefulness of a plane-wave basis in multiscale approaches formulated in real space. Over the past decade, numerous efforts have been directed towards the development of real-space DFT implementations that overcome these difficulties (e. g., Hehre et al. (1969); Wills and Cooper (1987); Soler et al. (2002); Skylaris et al. (2005); Bowler et al. (2006); Chelikowsky et al. (1994); Castro et al. (2006)). Though some of these studies use

the finite-element basis (Pask et al. (1999); Tsuchida (2004)), they do not utilize the unstructured nature of the the finite-element method and the convergence of the finite-element approximation with numerical quadratures has not been rigorously justified. This is one of the motivations of the current work along with the desire to coarse-grain DFT.

In this chapter, we present a non-periodic, real-space, finite-element formulation and implementation of DFT. For definiteness, we develop the formulation for the LSDA exchange-correlation functional. However, the formulation is not restricted to a particular type of exchange-correlation functional and can be adapted to other commonly used functionals such as GGA. The approaches adopted for solving the Kohn-Sham problem can be broadly classified as ones that use the self-consistent field (SCF) method and others in which some form of constrained direct minimization is performed (Payne et al. (1992)). However, these two approaches are equivalent (Parr and Yang (1989)) and largely a matter of convenience, as shown in Section 2.2.1. Our method of solution utilizes both SCF and variational schemes to ensure convergence and achieve the fastest possible convergence rate. One of the most time-consuming parts of conventional real-space implementations is the evaluation of the non-local electrostatic interactions, which we overcome by computing the electrostatic potential directly. This direct computation of the electrostatic field has the effect of turning the original minimization problem into a saddle-point problem (Ismail-Beigi and Arias (2000)). The Lagrangian functional thus defined is subsequently discretized by means of finite-elements and numerical quadrature. We show in Section 2.2.2 that the saddle-point problem is mathematically well-posed by proving existence of solutions, and we prove the Γ-convergence of finite-element approximations with numerical quadratures thereof, in Sections 2.2.3 and 2.2.4. We also present a similar analysis for the pseudopotential approximation in Section 2.2.5.

For completeness, we discuss in some detail relevant aspects of the finite-element implementation that are specific to the present application, including matters of optimal-mesh design, solution procedures (Section 2.3) and the computation of forces (Appendix A). In Section 2.4, we verify the formulation and numerical implementation by means of selected examples ranging from single atoms to small clusters. The core electrons are eliminated through the use of the Evanescent-Core (Fiolhais et al. (1995)) and Troullier-Martins (Troullier and Martins (1991)) pseudopotentials. To test the accuracy of the method, we also perform all-electron calculations for selected atoms and molecules. We also assess matters of numerical performance of the implementation, including convergence rates, scaling with problem size and parallel scalability.

2.1 Formulation

The theoretical framework of DFT has its origins in the Hohenberg-Kohn theorems (Hohenberg and Kohn (1964)). The first theorem is related to uniqueness and states that the ground state expectation value of any observable is a unique functional of the ground state electron density. The second theorem sheds light on the variational structure of the problem and shows that the electron density that minimizes the total energy is the exact ground state density. The problem of finding the ground state energy and electron density is equivalent to the problem of minimizing the energy of a system of non-interacting electrons in a mean-field. The corresponding energy functional $\mathcal{E} : X \times \mathbb{R}^{SM} \rightarrow \mathbb{R}$, where X is a suitable space of solutions for the orthogonal

wavefunctions, is given by (Parr and Yang (1989); Finnis (2003))

$$\mathcal{E}(\Psi, \mathbf{R}) = T_{\mathrm{s}}(\rho_\alpha, \rho_\beta) + E_{\mathrm{xc}}(\rho_\alpha, \rho_\beta) + E_{\mathrm{H}}(\rho_\alpha + \rho_\beta) + E_{\mathrm{ext}}(\rho_\alpha + \rho_\beta, \mathbf{R}) + E_{\mathrm{zz}}(\mathbf{R})\,, \tag{2.1}$$

where

$$\rho_\alpha(\mathbf{x}) = \sum_{i=1}^{N_\alpha} \psi_{i\alpha}^*(\mathbf{x})\psi_{i\alpha}(\mathbf{x}) = \sum_{i=1}^{N_\alpha} |\psi_{i\alpha}(\mathbf{x})|^2\,, \tag{2.2}$$

$$\rho_\beta(\mathbf{x}) = \sum_{i=1}^{N_\beta} \psi_{i\beta}^*(\mathbf{x})\psi_{i\beta}(\mathbf{x}) = \sum_{i=1}^{N_\beta} |\psi_{i\beta}(\mathbf{x})|^2 \tag{2.3}$$

represent the electron densities with the spin component 'up' and 'down' respectively and $\rho(\mathbf{x}) = \rho_\alpha(\mathbf{x}) + \rho_\beta(\mathbf{x})$ is the total electron density. By $\Psi = \{\psi_{1\alpha}, \psi_{2\alpha}, \ldots, \psi_{N_\alpha\alpha}, \psi_{1\beta}, \psi_{2\beta}, \ldots, \psi_{N_\beta\beta}\}$, we denote the vector of wavefunctions and $\mathbf{R} \in \mathbb{R}^{SM}$ is the collection of all the nuclear positions $\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M\}$, $S = 3$ denotes the space dimension and $M \in \mathbb{N}$ is the number of nuclei. The wavefunctions are orthonormal, thereby satisfying the relation

$$\int \psi_{i\sigma}^*(\mathbf{x})\psi_{j\sigma}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \delta_{ij}\,, \qquad \sigma \in \{\alpha, \beta\}\,, \quad i, j = 1, 2, \ldots, N_\sigma\,, \tag{2.4}$$

and

$$N_\alpha = \int \rho_\alpha(\mathbf{x})\,\mathrm{d}\mathbf{x}\,, \qquad N_\beta = \int \rho_\beta(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{2.5}$$

represent the number of electrons with spin 'up' and 'down' respectively. Note that if the domain of integration is not specified, it refers to all of space $\mathbb{R}^S$. The first term

$$T_{\mathrm{s}}(\rho_\alpha, \rho_\beta) = -\frac{1}{2} \sum_{\sigma} \sum_{i=1}^{N_\sigma} \int \psi_{i\sigma}^*(\mathbf{x}) \nabla^2 \psi_{i\sigma}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \tag{2.6}$$

in Eqn. (2.1) is the kinetic energy of the non-interacting electrons. The terms

$$\begin{aligned} E_{\mathrm{H}}(\rho) &= \frac{1}{2} \int \int \frac{\rho(\mathbf{x}) \rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{x}' \,, & (2.7) \\ E_{\mathrm{ext}}(\rho, \mathbf{R}) &= \int \rho(\mathbf{x}) V_{\mathrm{ext}}(\mathbf{x}, \mathbf{R}) \, \mathrm{d}\mathbf{x} \,, & (2.8) \\ E_{\mathrm{zz}}(\mathbf{R}) &= \frac{1}{2} \sum_{I=1}^{M} \sum_{\substack{J=1 \\ J \neq I}}^{M} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} & (2.9) \end{aligned}$$

are electrostatic terms with E_{H} known as the Hartree energy, representing the classical electrostatic interaction energy of the electron density, E_{ext} is the interaction energy with the external potential V_{ext} induced by the nuclear charges, and E_{zz} denotes the repulsive energy between the nuclei. The term $E_{\mathrm{xc}}(\rho_\alpha, \rho_\beta)$ in Eqn. (2.1) denotes the exchange-correlation energy. For definiteness, we specifically adopt the so-called LSDA in which the exchange-correlation energy is taken to be that of a uniform electron gas having the same local density (Kohn and Sham (1965)). The exchange-correlation energy can in turn be separated into individual contributions from the exchange and correlation parts, namely,

$$E_{\mathrm{xc}}(\rho_\alpha, \rho_\beta) = E_{\mathrm{x}}(\rho_\alpha, \rho_\beta) + E_{\mathrm{c}}(\rho_\alpha, \rho_\beta). \tag{2.10}$$

The expression for the exact exchange energy $E_{\mathrm{x}}(\rho_\alpha, \rho_\beta)$ of an uniform electron gas is

$$E_{\mathrm{x}}(\rho_\alpha, \rho_\beta) = -\frac{3}{4}\left(\frac{6}{\pi}\right)^{1/3} \int (\rho_\alpha^{4/3}(\mathbf{x}) + \rho_\beta^{4/3}(\mathbf{x}))\, \mathrm{d}\mathbf{x}\,. \tag{2.11}$$

In addition, the correlation energy $E_{\mathrm{c}}(\rho_\alpha, \rho_\beta)$ can be written as

$$E_{\mathrm{c}}(\rho_\alpha, \rho_\beta) = \int \varepsilon_{\mathrm{c}}(\rho_\alpha(\mathbf{x}), \rho_\beta(\mathbf{x}))\rho(\mathbf{x})\, \mathrm{d}\mathbf{x}. \tag{2.12}$$

Specifically, in applications we use the parametrization of Perdew and Wang (1992) fitted to accurate Monte Carlo simulations carried out by Ceperley and Alder (1980).

As expressed in Eqns. (2.7) and (2.9), the electrostatic interaction energy and the repulsive energy of the nuclei are non-local in nature and, thus are not amenable to a local discretization. In order to overcome this difficulty, we employ the following strategy. We begin by representing the nuclear charge of magnitude Z_I at a site $R_I \in \mathbb{R}^S$ by means of a regularized bounded charge distribution $-Z_I\delta_{R_I}(\mathbf{x})$ with a compact support in a neighborhood of a small ball around R_I (it is conventional in electronic-structure calculations to associate a negative charge with nuclei and a positive charge with electrons) and such that $\int_{\mathbb{R}^S} \delta_{R_I}(\mathbf{x})\, \mathrm{d}\mathbf{x} = 1$ for $1 \leq I \leq M$. We note that the electrostatic potential due to these regularized charge distributions is pointwise bounded. Defining as

$$b(\mathbf{x}, \mathbf{R}) = -\sum_{I=1}^{M} Z_I \delta_{R_I}(\mathbf{x}) \tag{2.13}$$

the sum of all such regularized charge distributions of the M nuclei present in the system, the

nuclear repulsive energy can be rewritten as

$$E_{\mathrm{zz}}(\mathbf{R}) = \frac{1}{2}\int\int \frac{b(\mathbf{x},\mathbf{R})b(\mathbf{x}',\mathbf{R})}{|\mathbf{x}-\mathbf{x}'|}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{x}'. \tag{2.14}$$

Notice that this differs from Eqn. (2.9) by the self-energy of the nuclei, but this is an inconsequential constant depending only on the nuclear charges. In addition, the electrostatic potential due to the nuclei and electron charge distribution can be computed as a solution to the Poisson equation

$$\frac{-1}{4\pi}\nabla^2\phi(\mathbf{x},\mathbf{R}) = \rho(\mathbf{x}) + b(\mathbf{x},\mathbf{R}). \tag{2.15}$$

This equation has the unique solution

$$\phi(\mathbf{x},\mathbf{R}) = \int \frac{\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}\,\mathrm{d}\mathbf{x}' + \int \frac{b(\mathbf{x}',\mathbf{R})}{|\mathbf{x}-\mathbf{x}'|}\,\mathrm{d}\mathbf{x}' = V_{\mathrm{H}}(\mathbf{x}) + V_{\mathrm{ext}}(\mathbf{x},\mathbf{R}). \tag{2.16}$$

Consequently, we have the variational problem

$$\begin{aligned}&\frac{1}{2}\int\int \frac{\rho(\mathbf{x})\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{x}' + \int \rho(\mathbf{x})V_{\mathrm{ext}}(\mathbf{x})\,\mathrm{d}\mathbf{x} + \frac{1}{2}\int\int \frac{b(\mathbf{x})b(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{x}'\\ &= -\inf_{\phi\in H_0^1(\mathbb{R}^S)}\left\{\frac{1}{8\pi}\int |\nabla\phi(\mathbf{x},\mathbf{R})|^2\mathrm{d}\mathbf{x} - \int(\rho(\mathbf{x})+b(\mathbf{x},\mathbf{R}))\phi(\mathbf{x},\mathbf{R})\,\mathrm{d}\mathbf{x}\right\}.\end{aligned} \tag{2.17}$$

Using this property, we can write Eqn. (2.1) as

$$\mathcal{E}(\Psi,\mathbf{R}) = \sup_{\phi\in H_0^1(\mathbb{R}^S)} \mathcal{L}(\Psi,\mathbf{R},\phi)\,, \tag{2.18}$$

where

$$\begin{aligned}\mathcal{L}(\Psi,\mathbf{R},\phi) &= -\frac{1}{2}\sum_{\sigma}\sum_{i=1}^{N_\sigma}\int \psi_{i\sigma}^*(\mathbf{x})\nabla^2\psi_{i\sigma}(\mathbf{x})\,\mathrm{d}\mathbf{x} + E_{\mathrm{xc}}(\rho_\alpha,\rho_\beta)\\ &\quad -\frac{1}{8\pi}\int |\nabla\phi(\mathbf{x},\mathbf{R})|^2\mathrm{d}\mathbf{x} + \int(\rho(\mathbf{x})+b(\mathbf{x},\mathbf{R}))\phi(\mathbf{x},\mathbf{R})\,\mathrm{d}\mathbf{x}. \qquad (2.19)\end{aligned}$$

Here and below, we denote by $H^m(\Omega)$ the space of m-times weakly differentiable functions in $L^2(\Omega)$ and by $H_0^m(\Omega) := \overline{C_0^\infty(\Omega)}^{\|\cdot\|_{H^m}}$, the space of Sobolev functions with zero boundary conditions in the trace sense.

The problem of determining the ground state electron density and the equilibrium positions of the nuclei can now be expressed as the saddle point problem

$$\inf_{\substack{\Psi\in(H_0^1(\mathbb{R}^S))^N\\ \mathbf{R}\in\mathbb{R}^{SM}}} \mathcal{E}(\Psi,\mathbf{R}) = \inf_{\substack{\Psi\in(H_0^1(\mathbb{R}^S))^N\\ \mathbf{R}\in\mathbb{R}^{SM}}} \sup_{\phi\in H_0^1(\mathbb{R}^S)} \mathcal{L}(\Psi,\mathbf{R},\phi) \qquad (2.20)$$

subject to the constraints

$$\int \psi_{i\sigma}^*(\mathbf{x})\psi_{j\sigma}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \delta_{ij}\,, \qquad \sigma\in\{\alpha,\beta\}, \quad i,j=1,2,\ldots,N_\sigma. \qquad (2.21)$$

In addition, N_α, N_β must be chosen such that the system has the lowest possible energy subject to the constraint of the total number of the electrons in the system being studied, i.e., $N_\alpha + N_\beta = N$.

2.2 Properties of the Kohn-Sham variational problem

In the present section, for the sake of clarity and notational simplicity, we will neglect spin polarization, thereby reducing the LSDA to the Local Density Approximation (LDA). This simplification is not essential, and the results presented in this section extend trivially to LSDA. Corresponding to Eqn. (2.1), for LDA we have

$$\mathcal{E}(\Psi, \mathbf{R}) = -\frac{1}{2}\sum_{i=1}^{N}\int \psi_i^*(\mathbf{x})\nabla^2\psi_i(\mathbf{x})\,\mathrm{d}\mathbf{x} + E_{\mathrm{xc}}(\rho) + E_{\mathrm{H}}(\rho) + E_{\mathrm{ext}}(\rho, \mathbf{R}) + E_{\mathrm{zz}}(\mathbf{R})\,, \tag{2.22}$$

where $\rho(\mathbf{x}) = \sum_{i=1}^{N}|\psi_i(\mathbf{x})|^2$ and $\Psi = \{\psi_1, \psi_2, \ldots, \psi_N\}$. Also, to avoid technical problems, we restrict our analysis to a bounded set $\Omega \subset \mathbb{R}^S$. With this simplification we obtain the variational problem

$$\inf_{\substack{\Psi\in(H_0^1(\Omega))^N \\ \mathbf{R}\in\mathbb{R}^{SM}}} \mathcal{E}(\Psi, \mathbf{R}) = \inf_{\substack{\Psi\in(H_0^1(\Omega))^N \\ \mathbf{R}\in\mathbb{R}^{SM}}} \sup_{\phi\in H_0^1(\Omega)} \mathcal{L}(\Psi, \mathbf{R}, \phi) \tag{2.23}$$

subject to the constraints

$$\int_\Omega \psi_i^*(\mathbf{x})\psi_j(\mathbf{x})\mathrm{d}\mathbf{x} = \delta_{ij}\,, \quad i, j = 1, 2, \ldots, N. \tag{2.24}$$

Note that in this section we do not fix S, the dimension of space.

The remainder of this section has been organized as follows. In Section 2.2.1, we show the equivalence of the Kohn-Sham variational and eigenvalue problems. The existence of a minimum for the Kohn-Sham variational principle is proved in Section 2.2.2 and in Sections 2.2.3, 2.2.4 we val-

idate the convergence of the finite-element approximation including numerical quadratures using the technique of Γ-convergence. Γ-convergence is an ideal and flexible tool for handling nonlinear variational problems like the one described by Eqns. (2.23), (2.24). Since a limitation of this method is its inability to provide the rates of convergence, we obtain it numerically in Section 2.4.4. The nonlinearity of the problem prevents the application of any of the standard techniques which provide the rate of convergence (Ciarlet (2002)).

2.2.1 Equivalence of variational and eigenvalue problems

We begin by showing the equivalence of the variational and eigenvalue problems in this section. The proof is similar in spirit to Roothaan (1951), where the equivalence was proved for the Hartree-Fock equations, and is provided here for the sake of completeness. Using Lagrange multipliers λ_{ij} to enforce the constraints given by Eqn. (2.24), we obtain the functional

$$\mathcal{E}_c(\Psi, \mathbf{R}, \Lambda) = \mathcal{E}(\Psi, \mathbf{R}) - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_{ij} \left(\int_\Omega \psi_i^*(\mathbf{x})\psi_j(\mathbf{x})\, \mathrm{d}\mathbf{x} - \delta_{ij} \right) . \tag{2.25}$$

Let us first prove that the matrix $\Lambda = (\Lambda_{ij})_{1\leq i,j\leq N}$, whose entries are λ_{ij}, is Hermitian. Taking variations of Eqn. (2.25) and setting them to zero we obtain

$$\left(-\frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{x}, \mathbf{R})\right)\psi_i(\mathbf{x}) = \sum_{j=1}^{N} \lambda_{ij}\psi_j(\mathbf{x}) , \tag{2.26}$$

$$\left(-\frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{x}, \mathbf{R})\right)\psi_i^*(\mathbf{x}) = \sum_{j=1}^{N} \lambda_{ji}\psi_j^*(\mathbf{x}) , \tag{2.27}$$

$$\int_\Omega \psi_i^*(\mathbf{x})\psi_j(\mathbf{x})\, \mathrm{d}\mathbf{x} = \delta_{ij} , \tag{2.28}$$

where

$$V_{\text{eff}}(\mathbf{x},\mathbf{R}) = V_{\text{ext}}(\mathbf{x},\mathbf{R}) + V_{\text{H}}(\mathbf{x}) + \frac{\delta E_{\text{xc}}(\rho)}{\delta\rho(\mathbf{x})}. \tag{2.29}$$

Taking the complex conjugate of Eqn. (2.27), then subtracting it from Eqn. (2.26), we obtain the relation

$$\sum_{j=1}^{N}(\lambda_{ij} - \lambda_{ji}^{*})\psi_j(\mathbf{x}) = 0. \tag{2.30}$$

Since the ψ_j are linearly independent it follows that $\lambda_{ij} = \lambda_{ji}^{*}$, showing the matrix Λ to be indeed Hermitian. For notational convenience, we may express Eqn. (2.26) in matrix form as

$$\mathcal{H}\Psi = \Psi\Lambda\,, \qquad \mathcal{H} = -\frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{x},\mathbf{R}). \tag{2.31}$$

Now let us subject the wavefunctions to a unitary transformation and the matrix Λ to a similarity transformation, i. e.,

$$\hat{\Psi} = \Psi Q, \qquad \hat{\Lambda} = Q^{*}\Lambda Q \tag{2.32}$$

with $QQ^{*} = Q^{*}Q = \mathbf{I}$, $\mathbf{I}$ being the identity matrix. Since unitary transformations are norm-conserving, the electron density remains invariant:

$$\rho(\mathbf{x}) = \sum_{i=1}^{N}|\psi_i(\mathbf{x})|^2 = \sum_{i=1}^{N}|\hat{\psi}_i(\mathbf{x})|^2. \tag{2.33}$$

Using the above relations we get

$$\hat{\mathcal{H}}\hat{\Psi} = \hat{\Psi}\hat{\Lambda} \tag{2.34}$$

where $\mathcal{H} = \hat{\mathcal{H}}$. Hence the wavefunctions on unitary transformation satisfy equations of exactly the same form as the original ones prior to transformation. Since the matrix Λ is Hermitian, there exists an unitary matrix Q such that $\hat{\Lambda}$ is a real diagonal matrix. Therefore, without any loss of generality we can replace Eqn. (2.31) by the eigenvalue problem

$$\mathcal{H}\psi_i = \epsilon_i \psi_i \,, \quad i = 1, 2, \ldots \; N \tag{2.35}$$

which is said to be in *canonical form*. Thus, the SCF eigenvalue problem and the direct variational formulation of the problem are equivalent and the choice of one over other is strictly a matter of convenience.

2.2.2 Existence of a minimum

In this section, we establish the existence of a minimum for the Kohn-Sham variational principle. The main result is Theorem 4. A more general proof of the same is given by Anantharaman and Cances (2008), where they do not make the assumption that Ω is bounded. However, our analysis differs from theirs and is necessary as it lays the groundwork for proving the convergence of the finite-element approximation including numerical quadrature in the subsequent sections.

We introduce a suitable space of solutions X corresponding to normalized orthogonal wavefunctions as

$$X = \left\{ \tilde{\Psi} \in (H_0^1(\Omega))^N \;\middle|\; \langle \tilde{\psi_i}, \tilde{\psi_j} \rangle_{(L^2(\Omega), L^2(\Omega))} = \delta_{ij}, \;\; i, j = 1, 2, \ldots, N \right\}. \tag{2.36}$$

We require that Ω be an open, bounded subset of $\mathbb{R}^S$ with Lipschitz boundary. We point out that if we additionally postulate that Ω is a dyadic cube, then all the subsequent results of this analysis also hold for

$$X = \left\{ \tilde{\Psi} \in (H^1_{\text{per}}(\Omega))^N \;\middle|\; \langle \tilde{\psi}_i, \tilde{\psi}_j \rangle_{(L^2(\Omega), L^2(\Omega))} = \delta_{ij}, \;\; i, j = 1, 2, \ldots, N \right\}, \tag{2.37}$$

which corresponds to a periodic system. By $H^1_{\text{per}}(\Omega)$ we denote the space of Sobolev functions on the torus obtained by identifying the opposite sides of Ω.

On integrating by parts, the energy of the system as expressed by Eqn. (2.22) can be rewritten in a generic form as

$$\mathcal{E}(\Psi, \mathbf{R}) = \sum_{i=1}^{N} \frac{1}{2} \int_\Omega |\nabla \psi_i(\mathbf{x})|^2 \mathrm{d}\mathbf{x} + J(\rho, \mathbf{R}) + G(\rho)\,, \qquad \rho(\mathbf{x}) = \sum_{i=1}^{N} |\psi_i(\mathbf{x})|^2, \tag{2.38a}$$

$$J(\rho, \mathbf{R}) = - \min_{\phi \in H^1_0(\Omega)} \left\{ \frac{1}{C_S} \int_\Omega |\nabla \phi(\mathbf{x}, \mathbf{R})|^2 \mathrm{d}\mathbf{x} - \int_\Omega (\rho(\mathbf{x}) + b(\mathbf{x}, \mathbf{R})) \phi(\mathbf{x}, \mathbf{R}) \, \mathrm{d}\mathbf{x} \right\} + \Sigma\,, \tag{2.38b}$$

$$G(\rho) = \int_\Omega g(\rho) \, \mathrm{d}\mathbf{x}\,, \tag{2.38c}$$

where C_S is a constant dependent on the dimension of space S and the term Σ specifies the self energy of the nuclear charges, which is an inconsequential constant in this analysis and hence will be dropped later.

We make the following hypothesis on g:

(A1) The density g is continuous in $\mathbb{R}^+$.

(A2) The growth condition $|g(t)| \leq c_2 |t|^q + c_1$ holds for positive real constants c_2, c_1 and the

exponent satisfies

$$q \in \begin{cases} \left[1, \frac{S}{S-2}\right), & \text{if } S > 2, \\ [1, +\infty), & \text{otherwise.} \end{cases}$$

(A3) The exponent q satisfies the condition $2q < 3$.

The need for Assumption (A3) becomes clear in the proof of Lemma 3.

Let the dual exponent p^* of p be given by $1/p^* = 1/p - 1/S$. We first note:

(a) For $\rho \in L^2(\Omega)$ the functional $J(\rho, \mathbf{R})$ is well defined. This result follows from the Poincaré inequality and the Lax-Milgram lemma. It is also straightforward to check that $\rho \in L^2(\Omega)$ if $2^* > 4$.

(b) The functional J is continuous in $L^2(\Omega)$.

(c) The functional G is continuous in $L^q(\Omega)$, which follows from the continuity of g and the growth condition (A2).

(d) As $\sum_{i=1}^N \frac{1}{2} \int_\Omega |\nabla \psi_i(\mathbf{x})|^2 \mathrm{d}\mathbf{x}$ is continuous in the strong topology of $(H_0^1(\Omega))^N$ and convex, it follows that $\Psi \mapsto \sum_{i=1}^N \frac{1}{2} \int_\Omega |\nabla \psi_i(\mathbf{x})|^2 \mathrm{d}\mathbf{x}$ is lower semi-continuous in the weak topology of $(H_0^1(\Omega))^N$.

Lemma 1 *X is closed in the weak topology of $(H_0^1(\Omega))^N$.*

Proof. Consider an arbitrary sequence $(\Psi_l)_l \subset X$ with $\Psi_l \rightharpoonup \Psi$ in $(H_0^1(\Omega))^N$. By the Rellich-Kondrakov theorem (Rudin (1991)), $H_0^1(\Omega)$ has a compact embedding into $L^2(\Omega)$. Thus $(\psi_{i,l})_l \to \psi_i$ in $L^2(\Omega)$ for $i = 1, 2, \ldots, N$ from which it follows that $\delta_{ij} = \langle \psi_{i,l} | \psi_{j,l} \rangle_{(L^2(\Omega), L^2(\Omega))} \to \langle \psi_i | \psi_j \rangle_{(L^2(\Omega), L^2(\Omega))}$

for $i, j = 1, 2, \ldots, N$ as $l \to \infty$. Therefore $\Psi \in X$, which implies that X is closed in the weak topology of $(H_0^1(\Omega))^N$.

Lemma 2 *Let (A1), (A2) hold and $S < 4$. Then $\mathcal{E}$ is lower semi-continuous in the weak topology of X.*

Proof. Consider any sequence $(\Psi_l)_l \subset (H_0^1(\Omega))^N$ such that $\Psi_l \rightharpoonup \Psi$ in $(H_0^1(\Omega))^N$. As by the Sobolev theorem the embedding $H_0^1(\Omega) \hookrightarrow L^4(\Omega)$ is compact for $S < 4$, it follows that $(\psi_{i,l})_l \to \psi_i$ in $L^4(\Omega)$ for $i = 1, 2, \ldots, N$. Hence $(\psi_{i,l}^2)_l \to \psi_i^2$ in $L^2(\Omega)$ and therefore $(\rho_l)_l \to \rho$ in $L^2(\Omega)$. From (A2) it follows that the embedding $H_0^1(\Omega) \hookrightarrow L^{2q}(\Omega)$ is compact. Thus $\rho_l \to \rho$ in $L^{\max\{2,q\}}(\Omega)$. From (b)–(d) it follows that $\sum_{i=1}^N \frac{1}{2} \int_\Omega |\nabla \psi_i(\mathbf{x})|^2 \mathrm{d}\mathbf{x}$, $J(\rho, \mathbf{R})$, and $G(\rho)$ are lower semi-continuous in the weak topology of $(H_0^1(\Omega))^N$. Hence, it follows that $\mathcal{E}$ is lower semi-continuous in the weak topology of $(H_0^1(\Omega))^N$, and as $X \subset (H_0^1(\Omega))^N$, the claimed lower semi-continuity of $\mathcal{E}$ in the weak topology of X follows.

Lemma 3 *Let (A2), (A3) hold. Then $\mathcal{E}$ is coercive in the weak topology of X.*

Proof. We first note the following results on the properties of $J(\rho, \mathbf{R})$:

(i) From the linearity of the Euler-Lagrange equations associated with $J(\rho, \mathbf{R})$, the electrostatic energy of the system can be rewritten as

$$J(\rho, \mathbf{R}) = E_H(\rho) + \int_\Omega V_{\mathrm{ext}}(\mathbf{x}, \mathbf{R}) \rho(\mathbf{x}) \mathrm{d}\mathbf{x} + L(\mathbf{R}), \tag{2.39a}$$

$$E_H(\rho) = -\min_{\phi \in H_0^1(\Omega)} \left\{ \frac{1}{C_S} \int_\Omega |\nabla \phi(\mathbf{x})|^2 \mathrm{d}\mathbf{x} - \int_\Omega \rho(\mathbf{x}) \phi(\mathbf{x}) \, \mathrm{d}\mathbf{x} \right\}, \tag{2.39b}$$

where $L(\mathbf{R}) = E_{zz}(\mathbf{R}) + \Sigma$.

We point out that V_{ext} is pointwise bounded. More precisely, it holds $|V_{\text{ext}}(\cdot, \mathbf{R})| \leq C$ almost everywhere in Ω for all $\mathbf{R} \in \mathbb{R}^{SM}$ and some constant C.

(ii) The functional E_H is super-linear, that is $E_H(\rho_1 + \rho_2) \geq E_H(\rho_1) + E_H(\rho_2)$ for arbitrary, almost everywhere positive ρ_1, $\rho_2 \in L^2(\Omega)$. This follows from the linearity of the corresponding Euler-Lagrange equations.

From (ii) we can directly derive a lower bound for $E_H(\rho)$,

$$\begin{aligned} E_H(\rho) &= E_H\Big(\sum_{i=1}^{N} |\psi_i|^2\Big) \geq \sum_{i=1}^{N} E_H(|\psi_i|^2) \\ &\geq \sum_{i=1}^{N} \left[\max_{\phi_i \in H_0^1(\Omega)} \left\{ \int_\Omega |\psi_i(\mathbf{x})|^2 \phi_i(\mathbf{x}) \mathrm{d}\mathbf{x} - \frac{1}{C_S} \int_\Omega |\nabla \phi_i(\mathbf{x})|^2 \mathrm{d}\mathbf{x} \right\} \right]. \end{aligned} \tag{2.40}$$

We use $\phi_i = C_0 |\psi_i|$ as test functions in Eqn. (2.40) for a constant C_0 that will be determined later. Additionally, we also recall the following simple result on the weak derivative of an absolute function, whose proof is elementary but can for instance be found in Gilbarg and Trudinger (1983). If $\psi_i \in H^1(\Omega)$, then $|\psi_i| \in H^1(\Omega)$ and $\|\nabla |\psi_i|\|_{L^2(\Omega)} \leq \|\nabla \psi_i\|_{L^2(\Omega)}$. Inserting $\phi_i(\mathbf{x}) = C_0 |\psi_i(\mathbf{x})|$ in Eqn. (2.40) as trial functions for $i = 1, 2, \ldots, N$, we arrive at the following lower bound for $E_H(\rho)$,

$$\begin{aligned} E_H(\rho) &\geq \sum_{i=1}^{N} \left[\max_{\phi_i \in H_0^1(\Omega)} \left\{ \int_\Omega |\psi_i(\mathbf{x})|^2 \phi_i(\mathbf{x}) \mathrm{d}\mathbf{x} - \frac{1}{C_S} \int_\Omega |\nabla \phi_i(\mathbf{x})|^2 \mathrm{d}\mathbf{x} \right\} \right] \\ &\geq \sum_{i=1}^{N} \left[C_0 \int_\Omega |\psi_i(\mathbf{x})|^3 \mathrm{d}\mathbf{x} - \frac{C_0}{C_S} \int_\Omega |\nabla \psi_i(\mathbf{x})|^2 \mathrm{d}\mathbf{x} \right] \\ &= C_0 \|\Psi\|_{L^3(\Omega)}^3 - \frac{C_0}{C_S} \|\nabla \Psi\|_{L^2(\Omega)}^2. \end{aligned} \tag{2.41}$$

Using the inequality given by Eqn. (2.41) in Eqn. (2.39a) and $\|V_{\text{ext}}\|_{L^\infty(\Omega)} \leq C$ we conclude

$$J(\rho) \geq C_0 \|\Psi\|^3_{L^3(\Omega)} - \frac{C_0}{C_S} \|\nabla\Psi\|^2_{L^2(\Omega)} - C\|\Psi\|^2_{L^2(\Omega)}. \tag{2.42}$$

Finally we require the inequality

$$G(\rho) \geq -c_0 \|\Psi\|^{2q}_{L^{2q}(\Omega)} - c_1. \tag{2.43}$$

This is a consequence of the growth condition (A2) on g which yields

$$G(\rho) \geq -c_2 \left\| \sum_{i=1}^{N} |\psi_i|^2 \right\|^q_{L^q(\Omega)} - c_1 \tag{2.44}$$

and a direct estimate of the norm in Eqn. (2.44).

With the help of Eqn. (2.42) and Eqn. (2.43) we end up with

$$\mathcal{E}(\Psi, \mathbf{R}) \geq \frac{1}{2}\left(1 - \frac{2C_0}{C_S}\right) \|\nabla\Psi\|^2_{L^2(\Omega)} + C_0\|\Psi\|^3_{L^3(\Omega)} - c_0\|\Psi\|^{2q}_{L^{2q}(\Omega)} - C\|\Psi\|^2_{L^2(\Omega)} - c_1. \tag{2.45}$$

Choosing $0 < 2C_0 < C_S$, we find that $\|\Psi\|^3_{L^3(\Omega)}$ grows faster than $\|\Psi\|^{2q}_{L^{2q}(\Omega)}$ as $2q < 3$ by (A3), and $\|\Psi\|^2_{L^2(\Omega)}$. This ensures $\mathcal{E}(\Psi) \to +\infty$ as $\|\Psi\|_{(H^1_0(\Omega))^N} \to +\infty$. This is the coercivity of $\mathcal{E}$ in the weak topology of $(H^1_0(\Omega))^N$. From Lemma 1, as X is closed in the weak topology of $(H^1_0(\Omega))^N$, it follows that $\mathcal{E}$ is coercive in the weak topology of X.

Theorem 4 *Let (A1)–(A3) hold and let $S < 4$. Then $\mathcal{E}$ possesses a minimizer in X.*

Proof. This is an immediate consequence of Lemma 2 and 3 and the fundamental theorem of the

calculus of variations, see, e. g., Dal Maso (1993); Struwe (1990).

The exchange and correlation functionals under LDA approximation satisfy the growth condition (A2) with $q = 4/3$. Hence, the results of this section apply for the energy functional associated with DFT. Theorem 4 establishes for $S \leq 3$ the existence of a self-consistent solution to the eigenvalue problem (Eqn. (2.35)), which determines the ground-state properties of a materials system.

2.2.3 Convergence of the finite-element approximation

Next we prove the convergence of finite-element approximation. We do so in two steps, first establishing the Γ-convergence of the restricted functional (Theorem 6), then its equi-coercivity (Lemma 7). The main result is Theorem 8. By $\mathcal{P}_k$ we denote the ring of polynomials of non-negative degree less than or equal to k for some fixed $k \geq 1$. Let T_h be a family of triangulations of Ω of decreasing mesh size $h > 0$, and let X_h be the corresponding sequence of subspaces of X consisting of functions whose restriction to every cell in T_h is a polynomial, i. e.,

$$X_h = \left\{ \tilde{\Psi} \in (H_0^1(\Omega))^N \;\middle|\; \langle \tilde{\psi}_i, \tilde{\psi}_j \rangle_{(L^2(\Omega), L^2(\Omega))} = \delta_{ij},\ i, j = 1, 2, \ldots, N,\ \tilde{\psi}_i | T \in \mathcal{P}_k \text{ for } T \in T_h \right\}.$$

Similarly, let

$$X_{1_h} = \left\{ \tilde{\phi} \in H_0^1(\Omega) \;\middle|\; \tilde{\phi} | T \in \mathcal{P}_k \text{ for } T \in T_h \right\} \tag{2.46}$$

be the corresponding family of subspaces X_{1_h} for the electrostatic problem. It follows from standard approximation theory that the sequence of spaces X_h and X_{1_h} become more and more dense

in X and $H_0^1(\Omega)$ as h decreases. We define a sequence of discrete energy functionals

$$\mathcal{E}_h(\Psi, \mathbf{R}) = \begin{cases} \frac{1}{2}\|\nabla\Psi\|^2_{L^2(\Omega)} + G(\rho) + J_h(\rho, \mathbf{R}), & \text{if } \Psi \in X_h, \\ +\infty, & \text{otherwise;} \end{cases}$$

where

$$J_h(\rho, \mathbf{R}) \;=\; -\min_{\phi \in H_0^1(\Omega)} I_h(\phi, \rho, \mathbf{R}),$$

and

$$I_h(\phi, \rho, \mathbf{R}) \;=\; \begin{cases} I(\phi, \rho, \mathbf{R}) & \text{if } \phi \in X_{1_h},\ \Psi \in X_h, \\ +\infty, & \text{otherwise,} \end{cases}$$

where

$$I(\phi, \rho, \mathbf{R}) = \frac{1}{C_S}\int_\Omega |\nabla\phi(\mathbf{x})|^2 \mathrm{d}\mathbf{x} - \int_\Omega (\rho(\mathbf{x}) + b(\mathbf{x}, \mathbf{R}))\phi(\mathbf{x}, \mathbf{R})\,\mathrm{d}\mathbf{x}. \tag{2.47}$$

The following remark, see Gavini et al. (2007b), is needed before proceeding to the proof of $\Gamma-$convergence of the finite-element approximation.

Remark 5 *If $(\Psi_h)_h \subset (X_h)$ is a sequence such that $\rho_h \to \rho$ in $L^2(\Omega)$, then*

$$\lim_{h\to 0} J_h(\rho_h) = J(\rho)$$

Theorem 6 *Let (A1), (A2) hold and let $S < 4$. Then $\mathcal{E}_h \to \mathcal{E}$ (in the Γ-sense) in the weak topology of X.*

Proof. To establish the $\Gamma-$convergence of the sequence of energy functionals we must establish the $\liminf$ inequality and construct a recovery sequence.

We first show the $\liminf$ inequality. Consider a sequence $(\Psi_h)_{h>0}$ such that $\Psi_h \rightharpoonup \Psi$ in X as $h \searrow 0$. We only have to discuss the case that there exists a subsequence $(\Psi_{h_k})_k$ with $\Psi_{h_k} \in X_{h_k}$ for every $k \in \mathbb{N}$. Otherwise the $\liminf$ inequality holds trivially,

$$+\infty = \liminf_{k\to\infty} \mathcal{E}_{h_k}(\Psi_{h_k}, \mathbf{R}) \geq \mathcal{E}(\Psi, \mathbf{R}). \tag{2.48}$$

By (A1), (A2) and since $S < 4$, it follows as in the proof of Lemma 2 that $\rho_{h_k} \to \rho$ in $L^{\max\{2,q\}}(\Omega)$ for $k \to \infty$. As $\mathcal{E}_h(\Psi, \mathbf{R}) \geq \frac{1}{2}\|\nabla\Psi\|^2_{L^2(\Omega)} + G(\rho) + J_h(\rho, \mathbf{R})$, we find

$$\liminf_{k\to\infty} \mathcal{E}_{h_k}(\Psi_{h_k}, \mathbf{R}) \geq \liminf_{k\to\infty} \left\{ \frac{1}{2}\|\nabla\ \Psi_{h_k}\|^2_{L^2(\Omega)} + G(\rho_{h_k}) + J_{h_k}(\rho_{h_k}, \mathbf{R}) \right\}. \tag{2.49}$$

From the compact embeddings of X into $L^2(\Omega)$ and $L^q(\Omega)$, Remark 5 which establishes the continuity of the discrete electrostatic problem in $L^2(\Omega)$, the continuity of G in $L^q(\Omega)$ from (c), and the lower semi-continuity of $\|\nabla\Psi_h\|^2_{L^2(\Omega)}$ in the weak topology of X from (d), it follows that

$$\liminf_{k\to\infty} \mathcal{E}_{h_k}(\Psi_{h_k}, \mathbf{R}) \geq \mathcal{E}(\Psi, \mathbf{R}). \tag{2.50}$$

This establishes the $\liminf$ inequality.

The construction of a recovery sequence is trivial from the density of the finite-element approximation spaces in $H^1(\Omega)$: Let $(\Psi_h)_{h>0}$ be a sequence constructed from the interpolation functions of successive triangulations such that $\Psi_h \to \Psi$ in X. From the continuity of individual terms of

the energy functional, including the discrete electrostatic interaction energy from Remark 5, we have $\lim_{h\to 0} \mathcal{E}_h(\Psi_h, \mathbf{R}) = \mathcal{E}(\Psi, \mathbf{R})$. Hence we conclude $\mathcal{E}_h \to \mathcal{E}$ in the weak topology of X.

Lemma 7 *Let the assumptions (A2), (A3) hold. Then the family $(\mathcal{E}_h)_{h>0}$ is equi-coercive in the weak topology of X.*

Proof. If $\Psi \notin X_h$, then $\mathcal{E}_h(\Psi, \mathbf{R}) = +\infty$. If $\Psi \in X_h$, then $\mathcal{E}_h(\Psi, \mathbf{R}) = \frac{1}{2}\|\nabla\Psi\|^2_{L^2(\Omega)} + G(\rho) + J_h(\rho, \mathbf{R})$. Also, as the finite-element subspaces X_h and X_{1_h} are constructed from a single triangulation, this allows us to use each component of $\Psi \in X_h$ as a trial function in the electrostatic problem as demonstrated in Lemma 3 and Eqn. (2.42). Consequently, we arrive at the inequality

$$J_h(\rho) \geq C_0\|\Psi\|^3_{L^3(\Omega)} - \frac{C_0}{C_S}\|\nabla\Psi\|^2_{L^2(\Omega)} - C\|\Psi\|^2_{L^2(\Omega)} \tag{2.51}$$

for arbitrary $C_0 > 0$. Using the growth results for G, we have uniformly in h

$$\mathcal{E}_h(\Psi) \geq \frac{1}{2}\left(1 - \frac{2C_0}{C_S}\right)\|\nabla\Psi\|^2_{L^2(\Omega)} + C_0\|\Psi\|^3_{L^3(\Omega)} - c_0\|\Psi\|^{2q}_{L^{2q}(\Omega)} - C\|\Psi\|^2_{L^2(\Omega)} - c_1. \tag{2.52}$$

Choosing $0 < 2C_0 < C_S$ and since $2q < 3$, the expression on the right-hand side is a coercive function independent of h in the weak topology of X. Thus it follows that $\mathcal{E}_h$ is equi-coercive in the weak topology of X.

Theorem 8 *Let the hypotheses (A1)–(A3) hold and let $S < 4$. Then*

$$\lim_{h\to 0} \inf_X \mathcal{E}_h = \min_X \mathcal{E}.$$

Proof. This result follows from Theorem 6, Lemma 7 and Dal Maso (1993, Theorem 7.8).

Remark 9 *Theorem 8 establishes rigorously the convergence of the ground state energy of a system computed with a finite-element approximation. Additionally, if the ground state is not degenerate, the convergence of the ground-state electron density in $H_0^1(\Omega)$ follows.*

2.2.4 Convergence of the finite-element approximation with numerical quadratures

An efficient implementation of the integrals in the functional requires a further approximation by numerical quadratures. In this section, we prove the convergence of the finite-element approximation with numerical quadratures. We do so by combining ideas introduced in the previous section with well-known estimates of quadrature errors (Ciarlet (2002)).

Let $I = \int_\Omega f(\mathbf{x})\,\mathrm{d}\mathbf{x}$ and if $\tilde{I}$ denotes a n-th order quadrature of I, the error associated with the numerical quadrature satisfies a bound of the type

$$|\tilde{I} - I| \leq C_\Omega^{n+1} \int_\Omega |D^{n+1} f(\mathbf{x})|\,\mathrm{d}\mathbf{x}, \tag{2.53}$$

where C_Ω is a Poincare-type constant related to the size of the domain. Let the degree of the polynomial used in the finite-element interpolation be k and identify h with the maximum element size in the mesh. We define the energy functional where all integrations are performed with numerical

quadrature rules (with $Q\{\mathcal{E}\} \equiv \tilde{\mathcal{E}}$ being the quadrature of $\mathcal{E}$) as

$$\tilde{\mathcal{E}}_h(\Psi, \mathbf{R}) = \begin{cases} Q\{\frac{1}{2}\|\nabla\Psi\|^2_{L^2(\Omega)}\} + \tilde{G}(\rho) + \tilde{J}_h(\rho, \mathbf{R}), & \text{if } \Psi \in X_h, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.54}$$

Here

$$\tilde{J}_h(\rho, \mathbf{R}) = -\min_{\phi \in H^1_0(\Omega)} \tilde{I}_h(\phi, \rho, \mathbf{R}),$$

and

$$\tilde{I}_h(\phi, \rho, \mathbf{R}) \;=\; \begin{cases} \tilde{I}(\phi, \rho, \mathbf{R}) = Q\{I(\phi, \rho, \mathbf{R})\}, & \text{if } \phi \in X_{1_h},\ \Psi \in X_h, \\ +\infty, & \text{otherwise.} \end{cases}$$

The error in the energy functional introduced by the numerical quadrature satisfies the estimate

$$\begin{aligned} |\tilde{\mathcal{E}}_h(\Psi, \mathbf{R}) - \mathcal{E}_h(\Psi, \mathbf{R})| \;\leq\; & Ch^{n+1} \sum_i \int_{e_i} \Big| D^{n+1} \Big[\sum_{j=1}^{N} |\nabla\psi_j(\mathbf{x})|^2 + g(\rho) \Big] \Big| \, \mathrm{d}\mathbf{x} \\ & + |\tilde{J}_h(\rho, \mathbf{R}) - J_h(\rho, \mathbf{R})|, \end{aligned}$$

where e_i denotes the i-th element in the finite-element mesh. For $\Psi \in X_h$, $|\nabla\psi_j(\mathbf{x})|$ is a polynomial function of degree $2(k-1)$ for $j = 1, 2, \ldots, N$. Thus, if $n - 2(k-1) \geq 0$, then $D^{n+1}[\sum_{j=1}^{N} |\nabla\psi_j|^2] = 0$. An application of the inverse inequality (Ciarlet (2002)) then gives

the alternative estimate

$$\begin{aligned} |\tilde{\mathcal{E}}_h(\Psi,\mathbf{R})-\mathcal{E}_h(\Psi,\mathbf{R})| &\leq Ch\sum_i\int_{e_i}\sum_{j=1}^{N}|2g'(\rho)\psi_j(\mathbf{x})\nabla\psi_j(\mathbf{x})|\,\mathrm{d}\mathbf{x}+|\tilde{J}_h(\rho,\mathbf{R})-J_h(\rho,\mathbf{R})| \\ &\leq Ch\|4(g'(\rho))^2\rho\|_{L^1(\Omega)}^{1/2}\|\nabla\Psi\|_{L^2(\Omega)}+|\tilde{J}_h(\rho,\mathbf{R})-J_h(\rho,\mathbf{R})|. \end{aligned} \quad (2.55)$$

Remark 10 *If $(\Psi_h)_h\subset(X_h)$ is a sequence such that $\Psi_h\rightharpoonup\Psi$ in X for $h\searrow 0$, if $n-2k+3>0$ and $S<4$, then $\lim_{h\to 0}\inf_{H_0^1(\Omega)}\tilde{I}_h(\cdot,\rho_h,\mathbf{R})=\min_{H_0^1(\Omega)}I(\cdot,\rho,\mathbf{R})$, i. e., $\lim_{h\to 0}\tilde{J}_h(\rho_h,\mathbf{R})=J(\rho,\mathbf{R})$ (Gavini et al. (2007b)).*

We denote by hypothesis (H) the following three assumptions:

(H.i) If $(\Psi_h)_h\subset(X_h)$ with $\Psi_h\rightharpoonup\Psi$ in X, then $\|(g'(\rho_h))^2\rho_h\|_{L^1(\Omega)}^{1/2}$ is bounded independently of h.

(H.ii) $S<4$.

(H.iii) $n-2k+3>0$.

Remark 11 *If hypotheses (H) hold, then it follows from Eqn. (2.55), Remark 5 and Remark 10 that $\lim_{h\to 0}\{\tilde{\mathcal{E}}_h(\Psi_h,\mathbf{R})-\mathcal{E}_h(\Psi_h,\mathbf{R})\}=0$.*

We now turn to the convergence of finite-element approximations with numerical quadratures.

Theorem 12 *Let (A1), (A2) and (H) hold. Then $\tilde{\mathcal{E}}_h$ Γ-converges to $\mathcal{E}$ as $h\to 0$ in the weak topology of X.*

Proof. Let $(\Psi_h)_h$ be a sequence such that $\Psi_h\rightharpoonup\Psi$. Without loss of generality we may assume the existence of a subsequence $(\Psi_{h_k})_k\subset(\Psi_h)_h$ with $\Psi_{h_k}\in X_{h_k}$ for every k, as otherwise it trivially

holds

$$+\infty = \liminf_{k\to\infty} \tilde{\mathcal{E}}_{h_k}(\Psi_{h_k}, \mathbf{R}) \geq \mathcal{E}(\Psi, \mathbf{R}). \tag{2.56}$$

We observe

$$\liminf_{k\to\infty} \tilde{\mathcal{E}}_{h_k}(\Psi_{h_k}, \mathbf{R}) \geq \liminf_{k\to\infty} \mathcal{E}_{h_k}(\Psi_{h_k}, \mathbf{R}) + \liminf_{k\to\infty} (\tilde{\mathcal{E}}_{h_k} - \mathcal{E}_{h_k})(\Psi_{h_k}, \mathbf{R}). \tag{2.57}$$

It follows from Remark 10 that $\lim_{k\to\infty}(\tilde{\mathcal{E}}_{h_k} - \mathcal{E}_{h_k})(\Psi_{h_k}, \mathbf{R}) = 0$. Consequently, from Theorem 6

$$\liminf_{k\to\infty} \tilde{\mathcal{E}}_{h_k}(\Psi_{h_k}, \mathbf{R}) \geq \liminf_{k\to\infty} \mathcal{E}_{h_k}(\Psi_{h_k}, \mathbf{R}) \geq \mathcal{E}(\Psi, \mathbf{R}). \tag{2.58}$$

This establishes the $\liminf$ inequality.

The construction of a recovery sequence is again trivial from the density of the finite-element approximation spaces in $H^1(\Omega)$. Let $(\Psi_h)_{h>0}$ be a sequence constructed from the interpolation functions of successive triangulations such that $\Psi_h \to \Psi$ in X. From Eqn. (2.55) we have

$$\lim_{h\to 0} \tilde{\mathcal{E}}_h(\Psi_h, \mathbf{R}) = \lim_{h\to 0} \mathcal{E}_h(\Psi_h, \mathbf{R}) = \mathcal{E}(\Psi, \mathbf{R}). \tag{2.59}$$

Hence, we conclude $\tilde{\mathcal{E}}_h \to \mathcal{E}$ in the weak topology of X.

Lemma 13 *If (A3) and (H) hold, then the energy functional $\tilde{\mathcal{E}}_h$ obtained by numerical quadrature is equi-coercive in the weak topology of X.*

Proof. We observe from Eqn. (2.45) that if $\Psi \in X_h$,

$$\mathcal{E}_h(\Psi, \mathbf{R}) \geq \frac{1}{2}\left(1 - \frac{2C_0}{C_S}\right)\|\nabla\Psi\|^2_{L^2(\Omega)} + C_0\|\Psi\|^3_{L^3(\Omega)} - c_0\|\Psi\|^{2q}_{L^{2q}(\Omega)} - C\|\Psi\|^2_{L^2(\Omega)} - c_1.$$

Since $2q < 3$, there exist positive constants C_1, C_2 and C_3 such that

$$\mathcal{E}_h(\Psi, \mathbf{R}) \geq C_1\|\nabla\Psi\|^2_{L^2(\Omega)} + C_2\|\Psi\|^2_{L^2(\Omega)} - C_3. \tag{2.60}$$

This implies

$$\tilde{\mathcal{E}}_h(\Psi, \mathbf{R}) \geq C_1\|\nabla\Psi\|^2_{L^2(\Omega)} + C_2\|\Psi\|^2_{L^2(\Omega)} - C_3 - Ch\|\Psi\|_{L^2(\Omega)}\|\nabla\Psi\|_{L^2(\Omega)}.$$

So there exists a bound $\overline{h} > 0$ such that for all $h < \overline{h}$

$$\mathcal{E}_h(\Psi, \mathbf{R}) \geq K_0\|\nabla\Psi\|^2_{L^2(\Omega)} + K_1\|\Psi\|^2_{L^2(\Omega)} - K_2, \tag{2.61}$$

where K_0, K_1, and K_2 are positive constants independent of h. From this inequality it follows that $\tilde{\mathcal{E}}_h$ is equi-coercive in the weak topology of X.

Theorem 14 *Let (A1)–(A3) and (H) hold. Then* $\lim_{h\to 0} \inf_X \tilde{\mathcal{E}}_h = \min_X \mathcal{E}$.

Proof. This result follows again from Theorem 12, Lemma 13, and Dal Maso (1993, Theorem 7.8).

For the LDA approximation of exchange-correlation functionals, hypothesis (A1)–(A3) hold

with $q = 4/3$. Also if $S \leq 3$, it is straightforward to check that (H.i) holds. Choosing appropriate quadrature rules which satisfy $n - 2k + 3 > 0$, the convergence of the self-consistent eigenvalue problem of DFT using a finite-element approximation with numerical quadratures is established.

2.2.5 Pseudopotential approximation

The tightly bound core electrons are chemically inactive and hence have a negligible contribution towards determining physical properties such as binding energies and bond lengths. The core states are localized in the vicinity of the nucleus leading to oscillations of the valence wavefunctions in this region due to the orthogonality requirement. Regardless of the basis set used, a large number of basis functions is needed to capture these oscillations. This difficulty may be overcome by recourse to the pseudopotential approximation (Pickett (1989)) in which the all-electron potential $V_{\text{ext}}(\mathbf{x}, \mathbf{R})$ in Eqn. (2.29) is replaced by an effective potential $V_{\text{ext}}^{\text{PS}}(\mathbf{x}, \mathbf{R})$ (or equivalently the energy $E_{\text{ext}}(\rho, \mathbf{R})$ in Eqn. (2.22) with an effective energy $E_{\text{ext}}^{\text{PS}}(\rho, \mathbf{R})$) and the core (non-valence) electrons are eliminated. The idea is that these effective potentials describe the effect of the core electrons adequately and represent the valence electrons by means of nodeless pseudo-wavefunctions. This approximation enables the solution of much larger systems than could otherwise be analyzed.

Based on their spatial dependence, pseudopotentials can be broadly classified as local and non-local. The non-local pseudopotentials, which include norm conserving (Bachelet et al. (1982); Rappe et al. (1990); Troullier and Martins (1991)) and ultrasoft pseudopotentials (Vanderbilt (1990)), are angular momentum dependent and are designed to accurately reproduce the scattering properties of the all-electron potential, thereby ensuring greater accuracy and transferability. They are

usually employed in the Kleinman-Bylander form (Kleinman and Bylander (1982)), which is less computationally expensive than the original semi-local form. In the sequel, we prove the existence of minimizers and convergence of the finite-element approximation with numerical quadratures for local as well as for non-local pseudopotential approximations.

2.2.5.1 Local pseudopotential

A local pseudopotential is an explicit function $V_{\text{ext}}^{\text{PS}}(\mathbf{x}, \mathbf{R})$ with $\|V_{\text{ext}}^{\text{PS}}\|_{L^\infty} \leq C$. The local pseudopotential approximation can then be incorporated into our variational formulation by replacing $b(\mathbf{x}, \mathbf{R})$ with $b^{\text{PS}}(\mathbf{x}, \mathbf{R}) = -\frac{2}{C_S}\nabla^2 V_{\text{ext}}^{\text{PS}}(\mathbf{x}, \mathbf{R})$. Consequently, all the results presented in the previous sections are applicable and hence existence of a minimum and convergence of the finite-element approximation with numerical quadratures follow.

2.2.5.2 Non-local pseudopotential

A non-local pseudopotential is an operator on the wavefunction ψ. In the Kleinman-Bylander form it is expressed for $J \in \{1, 2, \ldots, M\}$ as

$$V_{\text{ion}}^J(\mathbf{x}, \mathbf{R}_J)\psi(\mathbf{x}) = V_{\text{loc}}^J(\mathbf{x}, \mathbf{R}_J)\psi(\mathbf{x}) + \sum_{lm} C_{lm}^J u_{lm}^J(\mathbf{x}, \mathbf{R}_J)\Delta V_l{}^J(\mathbf{x}, \mathbf{R}_J), \tag{2.62}$$

where

$$\Delta V_l{}^J(\mathbf{x}, \mathbf{R}_J) = V_l{}^J(\mathbf{x}, \mathbf{R}_J) - V_{\text{loc}}^J(\mathbf{x}, \mathbf{R}_J)\,, \tag{2.63}$$

$$C_{lm}^J = \frac{\int_\Omega u_{lm}^J(\mathbf{x}, \mathbf{R}_J)\Delta V_l{}^J(\mathbf{x}, \mathbf{R}_J)\psi(\mathbf{x})\,\mathrm{d}\mathbf{x}}{\int_\Omega u_{lm}^J(\mathbf{x}, \mathbf{R}_J)\Delta V_l{}^J(\mathbf{x}, \mathbf{R}_J)u_{lm}^J(\mathbf{x}, \mathbf{R}_J)\,\mathrm{d}\mathbf{x}}\,. \tag{2.64}$$

$V_l^J(\mathbf{x}, \mathbf{R}_J)$ is the ionic pseudopotential component corresponding to the azimuthal quantum number l, $V_{\text{loc}}^J(\mathbf{x}, \mathbf{R}_J)$ is the local ionic potential and $u_{lm}^J(\mathbf{x}, \mathbf{R}_J)$ represents the pseudo-wavefunction for the valence states of interest, all for a single atom. The superscript J is for the atom number and the subscript m denotes the magnetic quantum number. Therefore

$$V_{\text{ext}}^{PS}(\mathbf{x}, \mathbf{R})\psi(\mathbf{x}) = \sum_{J=1}^{M} V_{\text{loc}}^J(\mathbf{x}, \mathbf{R}_J)\psi(\mathbf{x}) + \sum_{J=1}^{M}\sum_{lm} C_{lm}^J u_{lm}^J(\mathbf{x}, \mathbf{R}_J)\Delta V_l^J(\mathbf{x}, \mathbf{R}_J). \tag{2.65}$$

Let us redefine $V_{\text{ext}}(\mathbf{x}, \mathbf{R}) = \sum_{J=1}^{M} V_{\text{loc}}^J(\mathbf{x}, \mathbf{R}_J)$, for which we can obtain the corresponding $b(\mathbf{x}, \mathbf{R})$. The energy functional for the pseudopotential can therefore be written as

$$\mathcal{E}^{PS}(\Psi, \mathbf{R}) = \mathcal{E}(\Psi, \mathbf{R}) + \mathcal{K}(\Psi, \mathbf{R})\,, \tag{2.66}$$

where

$$\mathcal{K}(\Psi, \mathbf{R}) = \sum_{i=1}^{N_v}\sum_{J=1}^{M}\sum_{lm} \frac{1}{G_{lm}^J}\left|\int_\Omega f(\mathbf{x}, \mathbf{R}_J)\psi_i(\mathbf{x})\,\mathrm{d}\mathbf{x}\right|^2, \tag{2.67}$$

$$f(\mathbf{x}, \mathbf{R}_J) = u_{lm}^J(\mathbf{x}, \mathbf{R}_J)\Delta V_l^J(\mathbf{x}, \mathbf{R}_J)\,, \tag{2.68}$$

$$G_{lm}^J = \int_\Omega u_{lm}^J(\mathbf{x}, \mathbf{R}_J)\Delta V_l^J(\mathbf{x}, \mathbf{R}_J)u_{lm}^J(\mathbf{x}, \mathbf{R}_J)\,\mathrm{d}\mathbf{x}. \tag{2.69}$$

N_v is the total number of electrons after the pseudopotential approximation, i.e., valence electrons.

We make the assumption that $f \in H_0^1(\Omega)$ and $V_{\text{loc}}^J \in L^\infty(\Omega)$. Since V_{loc}^J is in principle arbitrary, it can be always chosen such that $G_{lm}^J \neq 0$. We also note the following properties:

(B1) The functional $\mathcal{K}$ is continuous in $L^2(\Omega)$.

This property can be directly verified using Hölder's inequality.

(B2) $\mathcal{K} \geq -C_1 \|\Psi\|^2_{L^2(\Omega)}$, where C_1 is a constant dependent on f, G^J_{lm}.

This inequality can immediately be obtained using Hölder's inequality.

Theorem 15 *Let (A1)–(A3) hold and let $S < 4$. Then $\mathcal{E}^{PS}(\Psi, \mathbf{R})$ possesses a minimizer in X.*

Proof. The lower semi-continuity of $\mathcal{E}^{PS}(\Psi, \mathbf{R})$ in the weak topology of X follows from Lemma 2 and (B1). From Lemma 3 and (B2) it is clear that $\mathcal{E}^{PS}(\Psi, \mathbf{R})$ is coercive in the weak topology of X. Therefore from the fundamental theorem of the calculus of variations, see, e. g., Dal Maso (1993), $\mathcal{E}^{PS}(\Psi, \mathbf{R})$ has a minimizer in X.

Theorem 16 *Let the hypotheses (A1)–(A3) hold and let $S < 4$. Then*

$$\lim_{h \to 0} \inf_X \mathcal{E}^{PS}_h = \min_X \mathcal{E}^{PS}.$$

Proof. Using the procedure outlined in Theorem 6 along with (B1) gives $\mathcal{E}^{PS}_h \to \mathcal{E}^{PS}$ (in the Γ-sense). The equi-coercivity of $(\mathcal{E}^{PS}_h)_{h>0}$ follows from Lemma 7 and (B2). Hence $\lim_{h\to 0} \inf_X \mathcal{E}^{PS}_h = \min_X \mathcal{E}^{PS}$ (Dal Maso (1993, Theorem 7.8)).

Theorem 17 *Let (A1)–(A3) and (H) hold. Then $\lim_{h\to 0} \inf_X \tilde{\mathcal{E}}^{PS}_h = \min_X \mathcal{E}^{PS}$.*

Proof. Consider $I = \int_\Omega f_1(x) f_2(x)\, \mathrm{d}\mathbf{x}$ for $f_1, f_2 \in H^1_0(\Omega)$. The error due to numerical quadrature is

$$\begin{aligned} |I - \tilde{I}| &\leq C_0 h^{n+1} \sum_i \int_{e_i} D^{n+1}(f_1 f_2)\, \mathrm{d}\mathbf{x} \\ &\leq Ch \left[\|\nabla f_1\|_{L^2(\Omega)} \|f_2\|_{L^2(\Omega)} + \|f_1\|_{L^2(\Omega)} \|\nabla f_2\|_{L^2(\Omega)} \right]. \end{aligned} \tag{2.70}$$

The last inequality is obtained by using the inverse inequality (Ciarlet (2002)) and Hölder's inequality. Therefore we have $\lim_{h\to 0}\{I-\tilde{I}\}=0$. Hence it follows from Remark 11 that $\lim_{h\to 0}\{\tilde{\mathcal{E}}_h^{PS}(\Psi_h,\mathbf{R})-\mathcal{E}_h^{PS}(\Psi_h,\mathbf{R})\}=0$. Adopting the same procedure as in Theorem 12 we obtain $\tilde{\mathcal{E}}_h^{PS}\to\mathcal{E}^{PS}$ (in the Γ-sense). Using (B2) in Lemma 13 gives the equi-coercivity of $\tilde{\mathcal{E}}_h^{PS}$ in the weak topology of X. From Dal Maso (1993, Theorem 7.8) $\lim_{h\to 0}\inf_X \tilde{\mathcal{E}}_h^{PS}=\min_X \mathcal{E}^{PS}$.

In this entire section, for notational simplicity we have not incorporated spin polarization into our analysis. Note however that this simplification has no bearing on the analysis/results presented here. In particular, it is clear that even with spin polarization, the required growth condition given by (A2) still holds, ensuring that all the results are applicable.

2.3 Numerical implementation

We now turn to the numerical implementation of the variational formulation described in Section 2.1. The variational problem (Eqn. (2.20)) is discretized using the finite-element interpolation scheme

$$\psi_{i\sigma}(\mathbf{x}) = \sum_{j=1}^{N_h}\psi_{i\sigma}^j N_j(\mathbf{x})\,, \tag{2.71}$$

$$\phi(\mathbf{x}) = \sum_{j=1}^{N_h}\phi^j N_j(\mathbf{x}) \tag{2.72}$$

to obtain

$$\sum_{j=1}^{N_h}\left[\int_\Omega\left(\frac{1}{2}\nabla N_j(\mathbf{x})\nabla N_k(\mathbf{x})+V^\sigma_{h,\text{eff}}(\mathbf{x},\mathbf{R})N_j(\mathbf{x})N_k(\mathbf{x})\right)\,\mathrm{d}\mathbf{x}\right]\psi^j_{i\sigma}$$
$$=\sum_{q=1}^{N_\sigma}\sum_{j=1}^{N_h}\lambda^\sigma_{iq}\psi^j_{q\sigma}\int_\Omega N_j(\mathbf{x})N_k(\mathbf{x})\,\mathrm{d}\mathbf{x}\,, \tag{2.73}$$

$$\sum_{j=1}^{N_h}\left[\frac{1}{4\pi}\int_\Omega\nabla N_k(\mathbf{x})\nabla N_j(\mathbf{x})\,\mathrm{d}\mathbf{x}\right]\phi^j=\int_\Omega(\rho_h(\mathbf{x})+b(\mathbf{x},\mathbf{R}))N_k(\mathbf{x})\,\mathrm{d}\mathbf{x}\,, \tag{2.74}$$

$$\sum_{s=1}^{N_h}\sum_{r=1}^{N_h}\left[\int_\Omega N_r(\mathbf{x})N_s(\mathbf{x})\,\mathrm{d}\mathbf{x}\right]\psi^r_{m\sigma}\psi^s_{n\sigma}=\delta_{mn}\,, \tag{2.75}$$

where $(N_k)_{1\leq k\leq N_h}$ are the basis functions of X_h, the Lagrange multipliers λ^σ_{iq} are used to enforce the constraints (Eqn. 2.21), $V^\sigma_{h,\text{eff}}$, ρ_h denote the discretized V^σ_{eff}, ρ respectively, $\sigma\in\{\alpha,\beta\}$ and $i,m,n=1,2,\ldots,N_\sigma$. In solving the nonlinear system of Eqns. (2.73), (2.74) and (2.75), the Newton-Raphson method suggests itself as a means of achieving quadratic convergence in the vicinity of the solution. However, the Newton-Raphson algorithm is only conditionally convergent and great care must be exercised to ensure that the initial guess lies within the radius of convergence of the solution. In addition, the quadratic convergence rate is only asymptotic and the method may exhibit slow convergence during the early stages of iteration. In view of these limitations of the Newton-Raphson method, an iterative solution scheme based on the equivalent generalized nonlinear eigenvalue problem

$$\mathbf{A}^\sigma\tilde{\Psi}^\sigma=\epsilon\mathbf{M}\tilde{\Psi}^\sigma,\qquad\sigma\in\{\alpha,\beta\} \tag{2.76}$$

may be advantageous early on in the iteration. Here,

$$\mathbf{A}^{\sigma}{}_{ij} = \frac{1}{2}\int_{\Omega} \nabla N_i(\mathbf{x}) \nabla N_j(\mathbf{x})\, \mathrm{d}\mathbf{x} + \int_{\Omega} V^{\sigma}_{\mathrm{h,eff}}(\mathbf{x}, \mathbf{R}) N_i(\mathbf{x}) N_j(\mathbf{x})\, \mathrm{d}\mathbf{x}\,, \tag{2.77}$$

$$\mathbf{M}_{ij} = \int_{\Omega} N_i(\mathbf{x}) N_j(\mathbf{x})\, \mathrm{d}\mathbf{x}, \tag{2.78}$$

and $\tilde{\Psi}$ is a vector of the nodal values of the wavefunction. Evidently, Eqn. (2.76) needs to be solved self-consistently with Eqn. (2.74). In general, a large numbers of finite-element basis functions per atom is required for accuracy and convergence (Pask and Sterne (2005)), thus rendering the SCF method computationally expensive since it now entails the repeated solution of large systems of linear equations.

Keeping the above discussion in mind, we employ the following methodology. Since we expect the electron density to decay much more rapidly than the electrostatic potential, the wavefunctions are solved on a smaller domain $\Omega^{\Psi} \subset \Omega$. We employ zero Dirichlet boundary conditions for both the wavefunctions and the electrostatic potential on their respective domains. We first construct a coarse Delaunay triangulation $\mathcal{T}_0$ of the domain Ω, with the nodes positioned at a coarsening rate of $r^{6/5}$ away from the nuclei, which is optimal for capturing the $1/r$ decay with linear interpolation. Such a triangulation has high resolution in Ω^{ψ} and coarsens away rapidly in $\Omega \setminus \Omega^{\psi}$, see Fig. 2.1 for a single-atom example. In addition, all nuclei are located on nodes of this triangulation.

A solution procedure that combines the best attributes of the Newton-Raphson and the SCF methods may be devised as follows. We start by solving Eqns. (2.74) and (2.76) using the SCF method on $\mathcal{T}_0$. Specifically, we employ the implicitly restarted Lanczos method (Saad (1992)) for solving the linear eigenvalue problem and the conjugate gradient algorithm for the linear solver.

We rewrite Eqn. (2.76) as

$$(\mathbf{A}^{\sigma} - \eta\mathbf{M})^{-1}\mathbf{M}\tilde{\Psi}^{\sigma} = \hat{\epsilon}\tilde{\Psi}^{\sigma}\,, \tag{2.79}$$

where η is the shift parameter and $\hat{\epsilon} = \frac{1}{\epsilon-\eta}$. By choosing an appropriate value for η, we magnify the part of the eigenvalue spectrum of interest, thereby significantly increasing the rate of convergence of the eigenvalue solver without incurring any additional cost. Also, since $\mathcal{T}_0$ is very coarse, the computational expense involved for this step is minimal. Next, we apply Freudenthal's tetrahedron subdivision algorithm (Bey and Aachen (2000)) to $\mathcal{T}_0$ to obtain a finer triangulation $\mathcal{T}_1$ (Fig. 2.2) that is an uniform subdivision of $\mathcal{T}_0$. The preliminary solution obtained on $\mathcal{T}_0$ is transferred to $\mathcal{T}_1$ using the shape functions on $\mathcal{T}_0$. This transferred solution serves as a starting guess for the solution of Eqns. (2.73), (2.74) and (2.75), simultaneously using Newton's method with Goldstein-Armijo line-searches (Dennis and Schnabel (1996)). For the associated linear solver we use the generalized minimal residual method (GMRES, Saad and Schultz (1986)) while retaining the option of switching to either the bi-conjugate gradient stable method (Bi-CGSTAB, van der Vorst (1992)) or transpose-free quasi-minimal residual method (TFQMR, Freund (1993)). Since the Hessian information is required only through its product with a vector, we evaluate it by using directional derivative quotients. This approach is significantly faster than evaluating the exact Hessian. However, we also use the exact Hessian when required. The quality of the initial guess obtained by the procedure just described starts the Newton-Raphson iteration well within the radius of convergence, thereby ensuring convergence within exceedingly tight tolerances at a quadratic rate in no more than 6 or 7 iterations.

The equilibrium position of the nuclei, defining an equilibrium configuration of the system, are

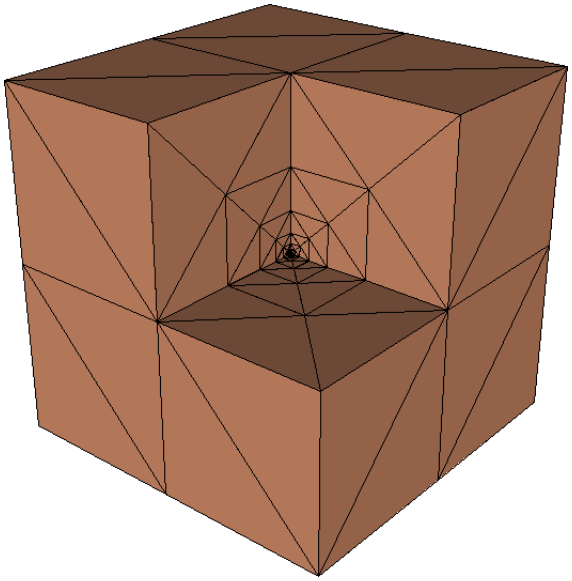


Figure 2.1: Mesh of a sliced cubical domain corresponding to the triangulation $\mathcal{T}_0$

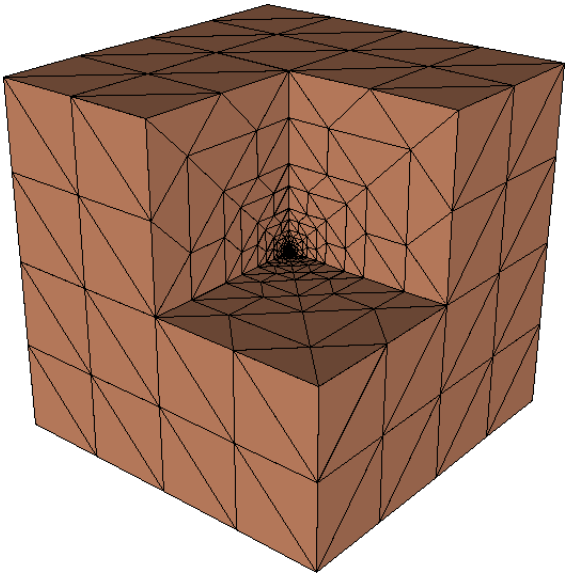


Figure 2.2: Mesh of a sliced cubical domain corresponding to the triangulation $\mathcal{T}_1$

computed using conjugate gradients with secant line search. The relevant forces to be equilibrated are derived in Appendix A. As the nuclei move, the triangulations must be updated. We do so simply by convecting the meshes according to the displacement field defined by the triangulation $\mathcal{T}_0$ on which the nuclei are located. The quality of these convected meshes is monitored throughout the calculations, and a complete remeshing is performed if the mesh quality falls below a prespecified tolerance. Each force update following a displacement of the nuclei requires the re-evaluation of the electrostatic potential and wavefunctions within an internal loop. In order to expedite this re-evaluation, we start from the initial electrostatic potential and wavefunctions that are convected from the previous configuration.

In order to verify convergence with respect to mesh size, we repeat the calculations on increasingly finer triangulations $\{\mathcal{T}_n\}$ obtained by successive uniform subdivisions. The solution on the parent triangulation is used as a starting guess on the finer triangulation, which ensures rapid convergence. This process is repeated until convergence is achieved to within a prespecified tolerance. The recursive nature of the calculations is advantageous with respect to solution schemes that require a complete restart every time the discretization is modified, e.g., plane wave basis calculations in which convergence has to be verified with respect to the energy cut-off and number of k-points sampled.

Finally, the formulation is implemented in parallel simply by recourse to domain decomposition. Owing to the local character of the majority of the calculations, domain decomposition may be expected to scale near-optimally. This expectation is indeed born out by numerical tests, see Section 2.4.4.

2.4 Examples and results

2.4.1 'All-electron' calculations

2.4.1.1 Atoms

The first set of examples considered are atoms of helium, lithium, carbon, nitrogen and oxygen. The triangulation $\mathcal{T}_0$ generated for these atoms, which has roughly 300 nodes, is depicted in Fig. 2.1. The convergence of the ground state energy of the helium atom on increasing the subdivision number (i.e., number of applications of the subdivision algorithm) is shown in Fig 2.3. It is clear that there is rapid convergence of the energy on decreasing the mesh size. This is representative of the other examples. Table 2.1 shows that we reproduce the ground state energies obtained by the highly accurate calculations of Kotochigova et al. (1997). The deviation from experiment reflects the well-known deficiencies of LSDA.

Table 2.1: Ground state energies of selected atoms (a.u)

Element	DFT-FE	KS-LSDA (Kotochigova et al. (1997))	Experiments (Veillard and Clementi (1968))
He	-2.833	-2.834	-2.904
Li	-7.340	-7.343	-7.478
C	-37.460	-37.470	-37.844
N	-54.125	-54.136	-54.587
O	-74.518	-74.527	-75.063

2.4.1.2 Molecules

The next set of examples are nitrogen (N_2) and carbon monoxide (CO) molecules. Tables 2.2 and 2.3 show that the binding energy and bond length that we obtain are in very good agreement with

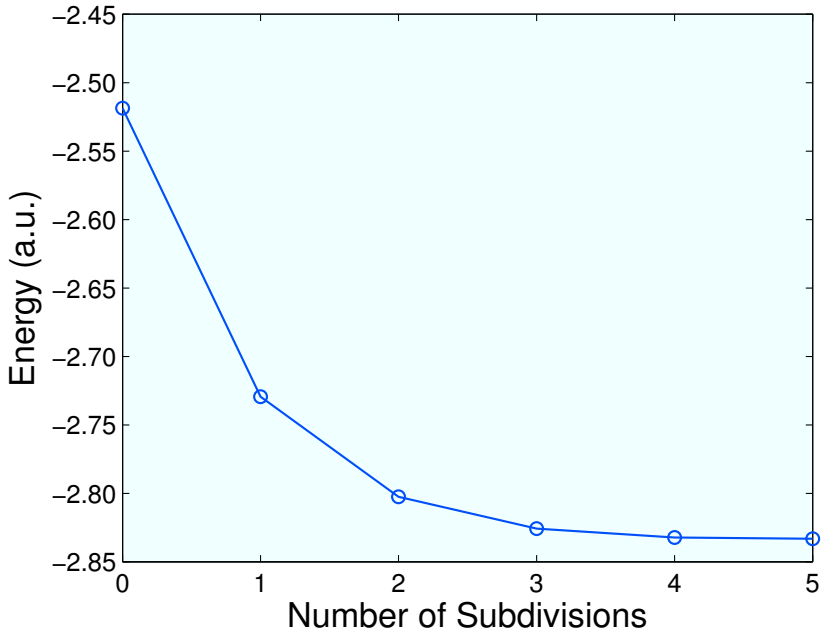


Figure 2.3: Energy of the helium atom as a function of the number of uniform subdivisions of triangulation $\mathcal{T}_0$

values in literature computed using plane waves (Engel et al. (2001)). Again, it is well known that LSDA predicts over-binding, i.e., higher binding energies and smaller bond lengths compared to experiments. The occupied valence molecular orbitals of CO can be seen in Fig. 2.4.

Table 2.2: Binding energy and bond length of N_2

Property	DFT-FE	KS-LSDA (Engel et al. (2001))	Experiments (Huber (1972))
Binding energy (eV)	-11.6	-11.593	-9.81
Bond length (a.u.)	2.06	2.068	2.07

Table 2.3: Binding energy and bond length of CO

Property	DFT-FE	KS-LSDA (Engel et al. (2001))	Experiments (Huber (1972))
Binding energy (eV)	-13.03	-12.967	-11.2
Bond length (a.u.)	2.08	2.128	2.13

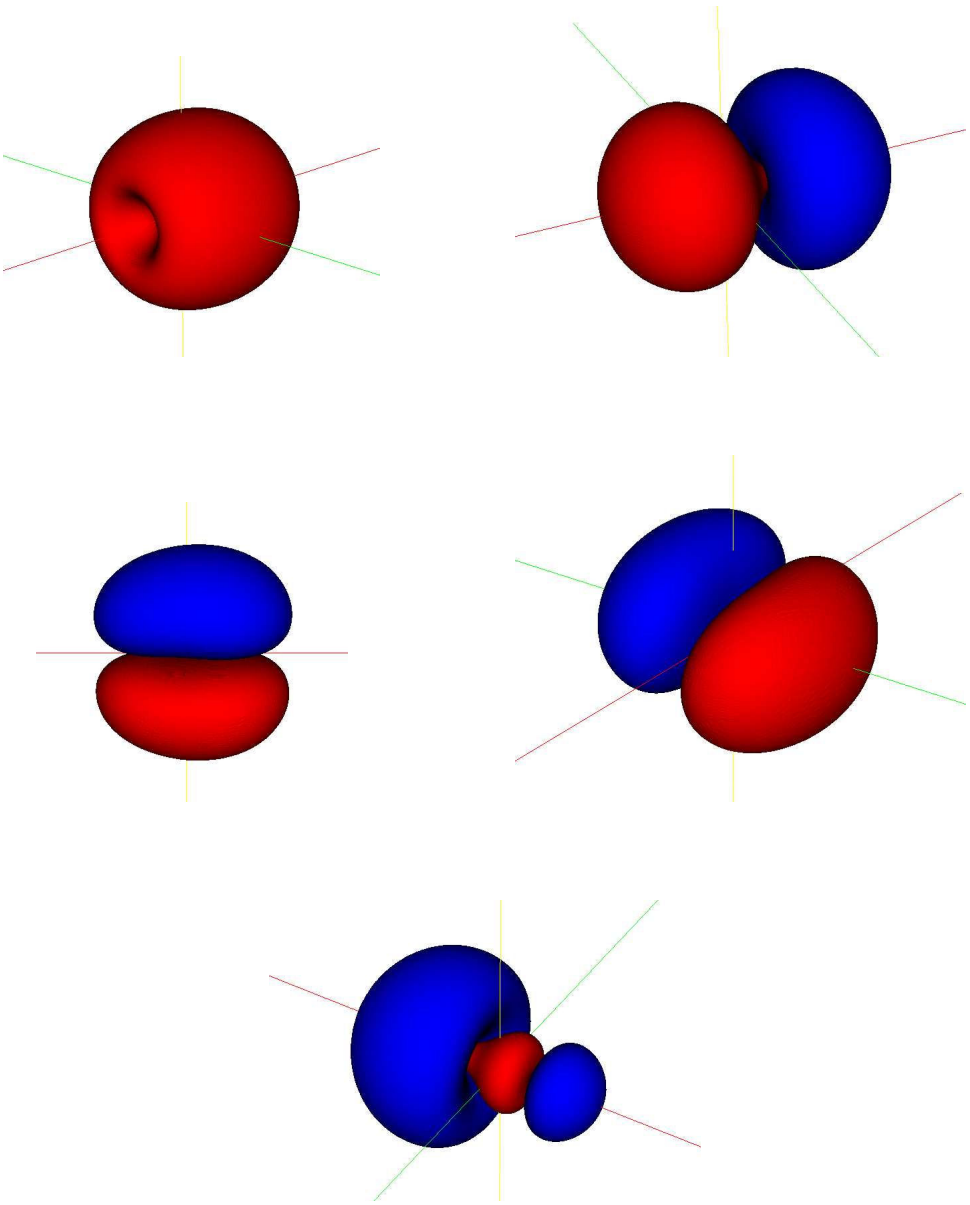


Figure 2.4: Occupied valence molecular orbitals of CO

2.4.2 Pseudopotential approximation (local)

In this section we present the results obtained using the smooth local 'Evanescent Core' pseudopotential (Fiolhais et al. (1995)). This pseudopotential has the form

$$V_{\text{ion}}^{I}(\mathbf{x}, \mathbf{R}_I) = -\frac{Z}{R_c}\left\{\frac{1}{y}(1-(1+\beta y)e^{-\alpha y}) - Ae^{-y}\right\}, \tag{2.80}$$

where Z is the number of valence electrons and $y = |\mathbf{x} - \mathbf{R}_I|/R_c$. The core decay length R_c and $\alpha \geq 0$ are element-dependent constants whose values can be obtained from Fiolhais et al. (1995). The relations used to evaluate β and A are

$$\beta = \frac{\alpha^3 - 2\alpha}{4(\alpha^2 - 1)}, \qquad A = \frac{1}{2}\alpha^2 - \alpha\beta. \tag{2.81}$$

The pseudopotential approximation is first used to calculate the pseudo-atom energy of lithium, sodium and magnesium. From Table 2.4, it is clear that we are able to replicate the results obtained by Nogueira et al. (1996). These pseudo-atom energies are further utilized to evaluate the binding energy and bond length of their respective dimers. As is evident from Table 2.5, the results obtained are in reasonable agreement with previous calculations (Nogueira et al. (1996)).

Table 2.4: Pseudo-atom energy (eV) using the 'Evanescent Core' pseudopotential

Metal	DFT-FE	Nogueira et al. (1996)
Li	-5.97	-5.97
Na	-5.21	-5.21
Mg	-23.05	-23.06

Next, we study the properties of some sodium clusters, in particular $1 \times 1 \times 1$, $2 \times 2 \times 2$ and

Table 2.5: Properties of selected metal dimers using 'Evanescent Core' pseudopotential

Dimer	Property	DFT-FE	Nogueira et al. (1996)
Li_2	Binding energy (eV)	-0.49	-0.52
	Bond length (a.u)	4.86	4.92
Na_2	Binding energy (eV)	-0.35	-0.46
	Bond length (a.u)	5.72	5.77
Mg_2	Binding energy (eV)	-0.06	-0.04
	Bond length (a.u)	7.12	7.18

$3 \times 3 \times 3$ body centered cubic (BCC) unit cells. A representative triangulation $\mathcal{T}_0$ used for $2 \times 2 \times 2$ BCC unit cells and its close-up view is shown in Figs. 2.5 and 2.6, respectively. We calculate the binding energy per atom and lattice constant for these clusters by computing the energy for various lattice distances and subsequently fitting the data to a cubic polynomial to obtain the point of minimum. The results so obtained are presented in Table 2.6. Also, the contours of electron density on the mid-plane and quarter-plane of $2 \times 2 \times 2$ BCC unit cells are shown in Figs. 2.7 and 2.8, respectively.

Table 2.6: Binding energy per atom and lattice constant of sodium BCC unit cells

Property	$1\times1\times1$	$2\times2\times2$	$3\times3\times3$
Binding energy/atom (eV)	-0.54	-0.70	-0.81
Lattice constant (a.u.)	7.20	7.55	7.75

2.4.3 Pseudopotential approximation (non-local)

Norm-conserving pseudopotentials are attractive because of their accuracy, transferability and availability for all elements in the periodic table. Here, we look at the results obtained using the Troullier-Martins (TM) pseudopotential (Troullier and Martins (1991)) implemented in the

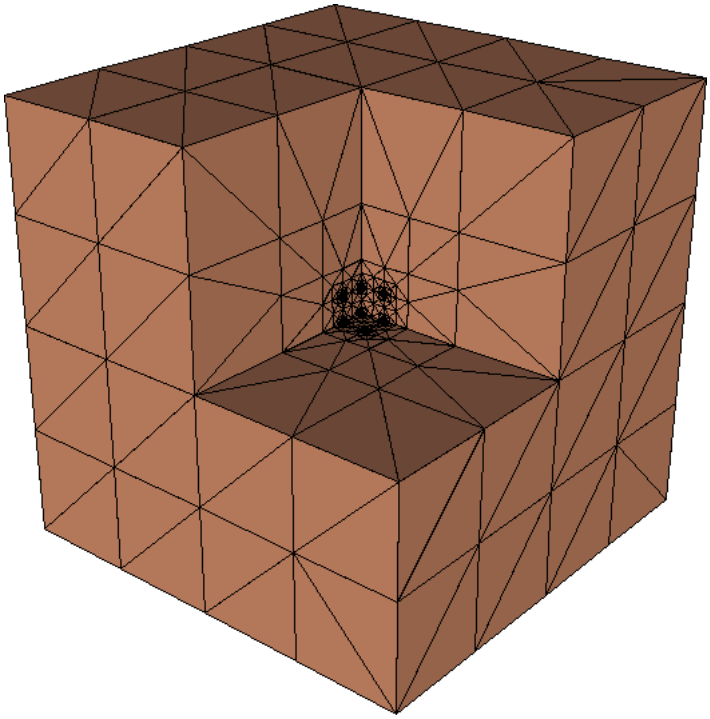


Figure 2.5: The triangulation $\mathcal{T}_0$ used for $2 \times 2 \times 2$ BCC unit cells

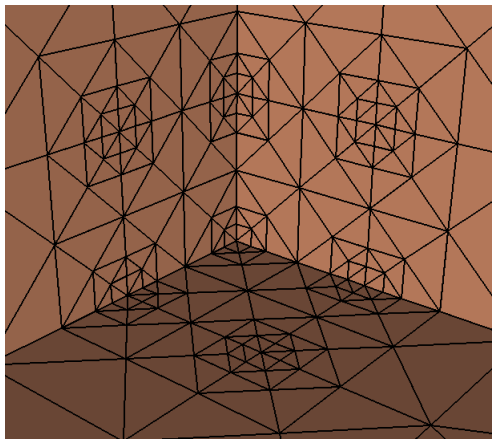


Figure 2.6: Close up of the triangulation $\mathcal{T}_0$ used for $2 \times 2 \times 2$ BCC unit cells

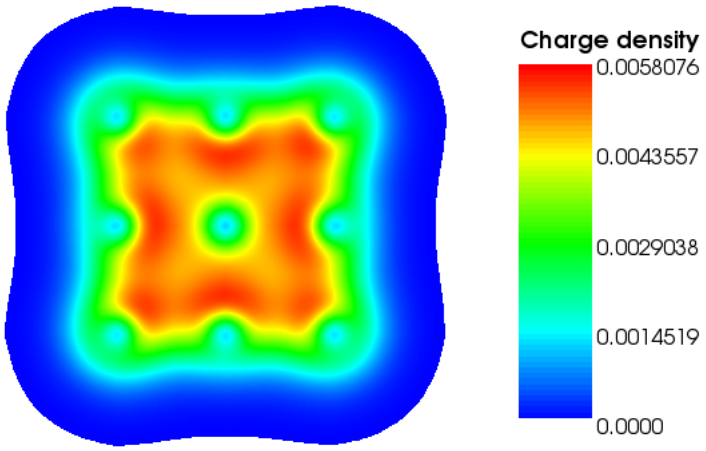


Figure 2.7: Contours of electron density on the mid-plane of a sodium cluster with 2×2×2 BCC unit cells

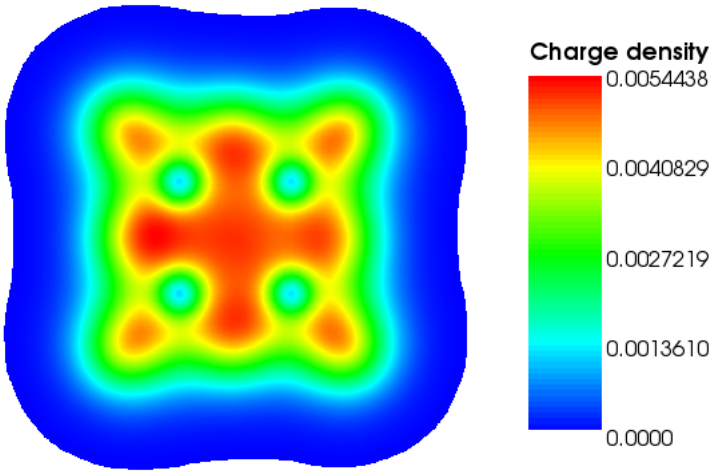


Figure 2.8: Contours of electron density on the quarter-plane of a sodium cluster with 2×2×2 BCC unit cells.

Kleinman-Bylander form.

The radial component of the TM pseudo-wavefunction for an atom located at $\mathbf{R}_I$ is given by

$$w_l(r) = \begin{cases} w_l^{AE}(r)\,, & r \geq r_c\,, \\ \\ r^l \exp[p(r)]\,, & r \leq r_c\,, \end{cases} \tag{2.82}$$

and the angular momentum dependent pseudopotential component has the form

$$V_l(r) = \begin{cases} V^{AE}(r)\,, & r \geq r_c\,, \\ \\ \epsilon_l + \frac{l+1}{r}\frac{p^{'}(r)}{2} + \frac{p^{''}(r)+(p^{'}(r))^2}{2}\,, & r \leq r_c. \end{cases} \tag{2.83}$$

The superscript AE stands for 'all-electron' calculation, ϵ_l represents the valence eigenvalues, $r = |\mathbf{x} - R_I|$ and r_c is the core radius. The coefficients of the polynomial $p(r) = c_0 + c_2 r^2 + c_4 r^4 + c_6 r^6 + c_8 r^8 + c_{10} r^{10} + c_{12} r^{12}$ are determined by norm conservation, continuity of the pseudo-wavefunction and its first four derivatives and the zero curvature of the pseudopotential at the origin.

In Tables 2.7, 2.8 we list the binding energy and bond length of B_2 and C_2 obtained using the TM pseudopotential and compare them with previous such studies (Engel et al. (2001)). The agreement is good, thereby validating the accuracy of our finite-element implementation.

Table 2.7: Binding energy and bond length of B_2

Property	DFT-FE	Engel et al. (2001)
Binding energy (eV)	-3.74	-3.79
Bond length (a.u.)	3.00	3.02

Table 2.8: Binding energy and bond length of C_2

Property	DFT-FE	Engel et al. (2001)
Binding energy (eV)	-6.86	-6.92
Bond length (a.u.)	2.35	2.35

2.4.4 Performance of the numerical method

This section is devoted to the assessment of the numerical performance of the implementation, specifically regarding the convergence rate with mesh size, scaling of the execution time with problem size and parallel scalability. We begin by assessing the rate of convergence of the finite-element method with decreasing mesh size. From Fig. 2.9, which depicts the normalized energy-error as a function of mesh size, we obtain $|\mathcal{E}^h - \mathcal{E}_0| \propto h^{1.9}$, where $\mathcal{E}_0$ is the converged value of the energy obtained by fitting the data to $\mathcal{E} = \mathcal{E}_0 + Ch^n$. This analysis thus shows that the energy converges quadratically, which is the expected rate of convergence for linear interpolation (Ciarlet (2002)).

Fig. 2.10 collects execution times as a function of the number of nodes in the triangulation for two selected examples. As is evident from the figure, the execution time scales linearly, or $O(N)$, asymptotically with the size N of the problem. This scaling is expected, since the complexity of the solution procedure is dominated by the complexity of the solution of a banded system of equations, which is $O(N)$ when the system bandwidth is independent of N. This latter property is typical of finite-elements, for which the system bandwidth is determined by the number of nearest neighbors of a typical node. Thus, our formulation achieves linear, or $O(N)$, scaling with respect to the number of nodes in the triangulation for a fixed number of electrons.

Finally, Fig. 2.11 illustrates the parallel efficiency of the implementation as measured by the

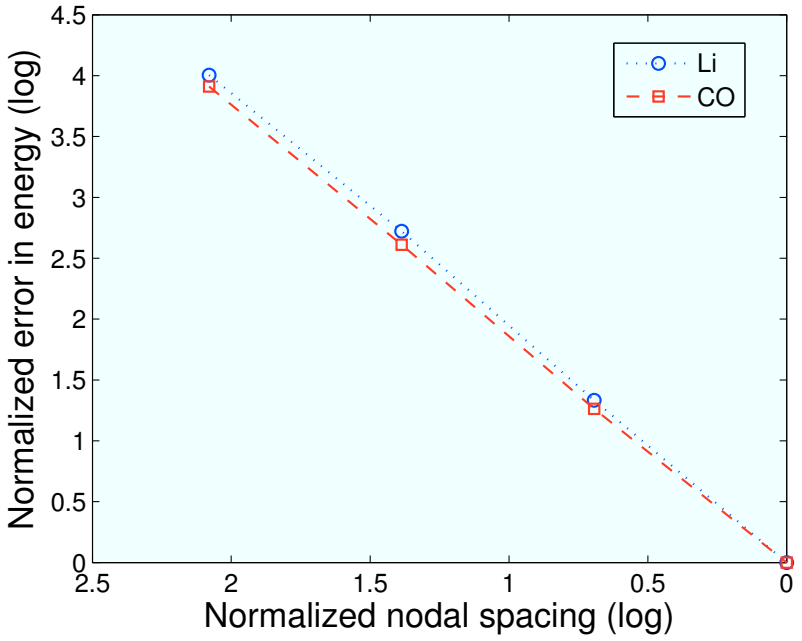


Figure 2.9: Convergence rate of the finite-element method

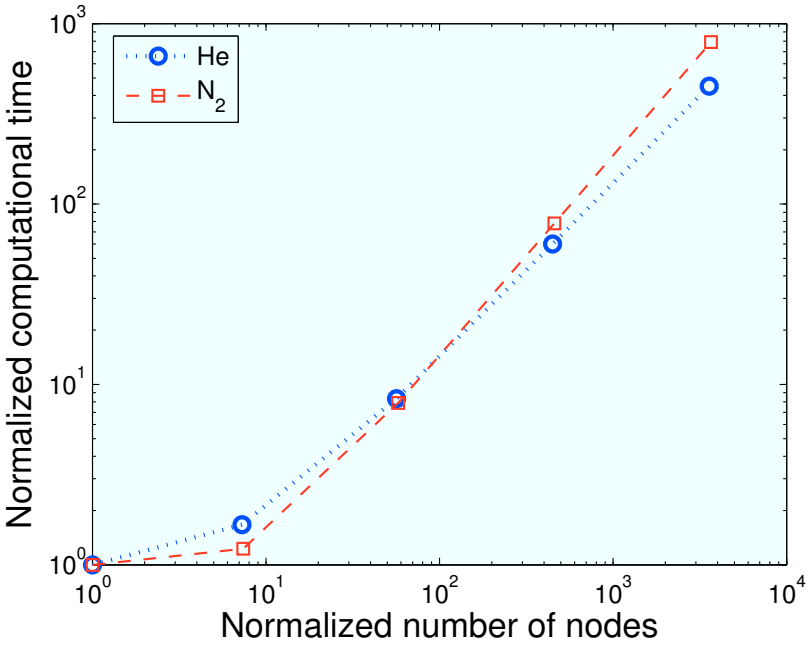


Figure 2.10: Scaling of computational time with number of nodes

relative speedup with increasing number of processors on a fixed problem of constant size. As is evident from the figure, the scaling is roughly linear. However, as the number of processors increases, the parallel efficiency correspondingly decreases owing to the additional communication required between them. Thus, the relative speedup corresponding to a 16-fold increase in the number of processors is of the order of 10.25. This speed-up factor translates into an efficiency of about 0.65, which is in the ballpark expected of domain decomposition applied to medium-size problems.

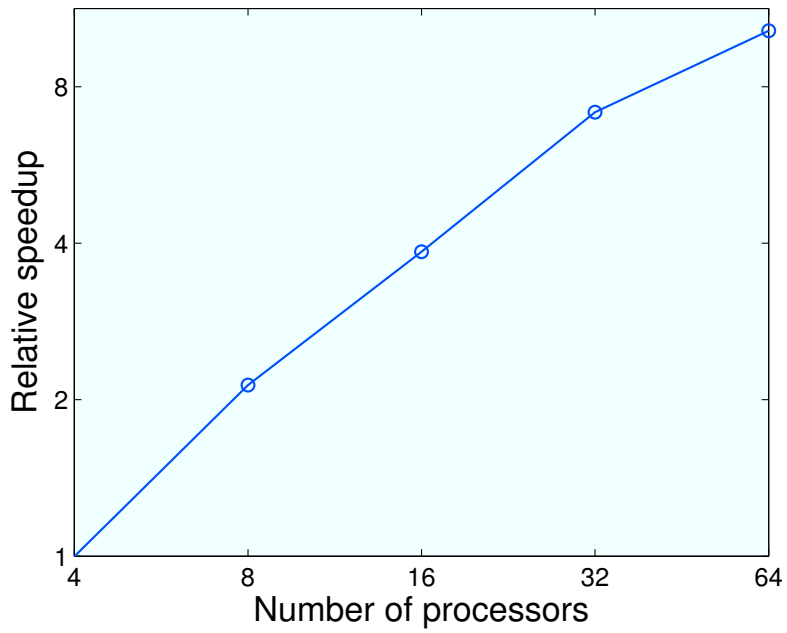


Figure 2.11: Relative speedup as a function of the number of processors

Chapter 3

Mesh-free convex approximation scheme for Kohn-Sham density functional theory

In the previous chapter, we provided a non-periodic, finite-element formulation for DFT. However, the finite-element method requires the generation of a mesh with particular care to nodal connectivity. Further, a large number of basis functions are required, rendering the method expensive for solving DFT. In this chapter, we develop a non-periodic, real space, mesh-free convex approximation scheme for DFT. We use *max-ent* basis functions which are the Pareto optimum between maximizing the entropy and the locality of the approximation scheme (Arroyo and Ortiz (2006)). The finite-element method can be recovered as a special case of the approximation scheme presented here (Arroyo and Ortiz (2006)). It provides advantages over conventional mesh-based methods in terms of the flexibility and adaptivity of the spatial discretization. Additionally, it is known to be significantly more accurate than the corresponding simplicial finite-elements (Arroyo and Ortiz (2006); Cyron et al. (2009)), also verified by the results obtained in this work. The limitations of the method in comparison to the finite-element method include the larger support-size of the basis functions and the need for expensive quadrature rules.

The outline of this chapter is as follows. In Section 3.1 we provide a brief introduction to the

convex approximation scheme. In Section 3.2, we discuss in some detail the numerical implementation, which we validate through select examples in Section 3.3.

3.1 Convex approximation scheme — *max-ent* basis functions

In this section, we introduce the convex approximation scheme based on *max-ent* basis functions (Arroyo and Ortiz (2006)). Consider a set of distinct nodes $Y = \{\mathbf{x}_a, a = 1, \ldots, P\} \subset \mathbb{R}^3$ whose convex hull is denoted by Ω. We are looking to create approximations to any function $u : \Omega \to \mathbb{R}$ of the form

$$u^h(\mathbf{x}) = \sum_{a=1}^{P} u_a p_a(\mathbf{x}) \,, \tag{3.1}$$

where $p_a : \Omega \to \mathbb{R}$ are the basis/interpolation functions. We require that the basis functions satisfy the following constraints $\forall \mathbf{x} \in \Omega$

$$\sum_{a=1}^{P} p_a(\mathbf{x}) = 1 \,, \tag{3.2}$$

$$\sum_{a=1}^{P} p_a(\mathbf{x})\mathbf{x}_a = \mathbf{x} \,, \tag{3.3}$$

$$p_a(\mathbf{x}) \geq 0 \,, \quad a = 1, \ldots, P. \tag{3.4}$$

Eqns. (3.2) and (3.3) ensure that we are able to exactly reproduce affine functions, akin to simplicial linear finite-element basis functions. The approximation scheme is called convex since it follows from Eqns. (3.2), (3.3) and (3.4), that the basis functions at $\mathbf{x} \in \Omega$ define a convex combination of nodes which evaluates to $\mathbf{x}$.

Let us define by

$$H(\mathbf{x},\mathbf{p}) = -\sum_{a=1}^{P} p_a(\mathbf{x})\log(p_a(\mathbf{x}))\,, \tag{3.5}$$

$$U(\mathbf{x},\mathbf{p}) = \int_{\Omega}\sum_{a=1}^{P} p_a(\mathbf{x})h_a(|\mathbf{x}-\mathbf{x}_a|)\,\mathrm{d}\mathbf{x}\,, \tag{3.6}$$

the information entropy and the total width of the convex approximation scheme, respectively. The vector of basis functions is given by $\mathbf{p}(\mathbf{x}) = \{p_1(\mathbf{x}),\ldots,p_P(\mathbf{x})\}$ and $h_a(|\mathbf{x}-\mathbf{x}_a|)$ denotes some generalized notion of distance between points $\mathbf{x}$, $\mathbf{x}_a$. Since Eqn. (3.6) does not involve any basis function derivatives, we can work with the pointwise width, using the functional

$$W(\mathbf{x},\mathbf{p}) = \sum_{a=1}^{P} p_a(\mathbf{x})h_a(|\mathbf{x}-\mathbf{x}_a|). \tag{3.7}$$

Ideally, we would want an approximation scheme that maximizes the entropy and minimizes the width. However, this is not always possible and therefore we look to find the Pareto optimum: a convex approximation scheme such that there is none better, by minimizing for fixed $\mathbf{x}$, $\beta \in (0,\infty)$

$$F(\mathbf{x},\mathbf{p}) = -H(\mathbf{x},\mathbf{p}) + \beta W(\mathbf{x},\mathbf{p}) \tag{3.8}$$

subject to the constraints (3.2), (3.3) and (3.4). Thus, we end up with the Lagrangian

$$\mathcal{L}(\mathbf{x},\mathbf{p},\lambda_0,\boldsymbol{\lambda}) = -H(\mathbf{x},\mathbf{p}) + \beta W(\mathbf{x},\mathbf{p}) + \lambda_0\left(\sum_{a=1}^{P} p_a - 1\right) + \boldsymbol{\lambda}.\sum_{a=1}^{P} p_a(\mathbf{x}_a - \mathbf{x}) \tag{3.9}$$

where $\lambda_0 \in \mathbb{R}$ and $\boldsymbol{\lambda} \in \mathbb{R}^3$ are Lagrange multipliers used to enforce the constraints given by Eqns. (3.2) and (3.3) respectively. As in Arroyo and Ortiz (2006), we differentiate between the interior points $\mathbf{x} \in \mathrm{int}(\Omega)$ and points on the boundary $\mathbf{x} \in \mathrm{bd}(\Omega)$.

3.1.1 Interior points

Definition 18 *The partition function $Z : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ associated with the node set Y is*

$$Z(\mathbf{x}, \boldsymbol{\lambda}) \equiv \sum_{a=1}^{P} \exp(-\beta h_a(|\mathbf{x} - \mathbf{x}_a|) + \boldsymbol{\lambda}.(\mathbf{x} - \mathbf{x}_a)). \tag{3.10}$$

Proposition 19 *Suppose* $\mathrm{affine}\, Y = \mathbb{R}^3$ *and* $\mathbf{x} \in \mathrm{int}(\Omega)$. *Then the solution of minimizing (3.8) subject to the contraints (3.2), (3.3) and (3.4) is*

$$p_a(\mathbf{x}) = \frac{1}{Z(\mathbf{x}, \boldsymbol{\lambda}^*(\mathbf{x}))} \exp(-\beta h_a(|\mathbf{x} - \mathbf{x}_a|) + \boldsymbol{\lambda}^*(\mathbf{x}).(\mathbf{x} - \mathbf{x}_a)), \quad a = 1, \ldots, P \tag{3.11}$$

where

$$\boldsymbol{\lambda}^*(\mathbf{x}) = \arg\min_{\boldsymbol{\lambda} \in \mathbb{R}^3} \log Z(\mathbf{x}, \boldsymbol{\lambda}). \tag{3.12}$$

Additionally, the minimizer $\boldsymbol{\lambda}^*(\mathbf{x})$ is unique.

Proof. *It follows from the Karush-Kuhn-Tucker conditions*

$$\beta h_a + \log(p_a) + 1 + \lambda_0^* + \boldsymbol{\lambda}^*.(\mathbf{x}_a - \mathbf{x}) = 0, \quad a = 1, \ldots, P. \tag{3.13}$$

Therefore

$$p_a = \exp(-\beta h_a + \boldsymbol{\lambda}^*.(\mathbf{x} - \mathbf{x}_a) - \lambda_0^* - 1) \quad \forall\, x \in \mathrm{int}(\Omega). \tag{3.14}$$

The optimal Lagrange multipliers λ_0^* *and* $\boldsymbol{\lambda}^*$ *are the maximizers of the Lagrange dual function*

$$g(\lambda_0, \boldsymbol{\lambda}) = \inf_{\mathbf{p} \in \mathbb{R}^N_+} \mathcal{L}(\mathbf{p}, \lambda_0, \boldsymbol{\lambda}) = -\lambda_0 - \sum_{a=1}^{P} \exp(-\beta h_a + \boldsymbol{\lambda}.(\mathbf{x} - \mathbf{x}_a) - \lambda_0 - 1) \tag{3.15}$$

On maximizing this dual function with respect to λ_0*, we obtain*

$$Z(\mathbf{x}, \boldsymbol{\lambda}) = \exp(\lambda_0^* + 1). \tag{3.16}$$

Substituting Eqn. (3.16) into (3.15), we obtain the reduced Lagrange dual function

$$\hat{g}(\boldsymbol{\lambda}) = -\log Z(\mathbf{x}, \boldsymbol{\lambda}). \tag{3.17}$$

from which we can conclude that if $\mathbf{x} \in \mathrm{int}(\Omega)$*, then the shape functions are given by Eqns. (3.11) and (3.12). The existence of minimizer of* $-\hat{g}(\boldsymbol{\lambda})$ *is guaranteed by the Kuhn-Tucker theorem. The uniqueness follows from the strict convexity of* $\log Z(\mathbf{x}, \boldsymbol{\lambda})$*, which can easily be verified.*

3.1.2 Boundary points

The treatment of boundary points $\mathbf{x} \in \mathrm{bd}(\Omega)$ can be reduced to the problem analyzed in the preceding section by exploiting the reduced face property of convex approximation schemes (Arroyo and Ortiz (2006)).

The adaptability of the basis functions stems from the flexibility provided by the choice of $h_a(|\mathbf{x}-\mathbf{x}_a|)$ and $\beta(\mathbf{x})$. Let us first discuss the effect of the parameter $\beta(\mathbf{x})$. The support size of the basis functions, which can also be optimized in a variational approach (Rosolen et al. (2010)), is determined by $\beta(\mathbf{x})$. Reducing β increases the support of the basis functions and in general the accuracy of the calculations (Arroyo and Ortiz (2006)). However, it comes at the cost of higher order quadrature rules and increased band-width of the equations. Alternatively, increasing β increases the locality of the basis functions at the price of reduced accuracy. Detailed examples and illustrations can be found elsewhere (Arroyo and Ortiz (2006)).

Further adaptivity is provided by the choice of $h_a(\mathbf{x}-\mathbf{x}_a)$. By choosing $h_a(|\mathbf{x}-\mathbf{x}_a|) = |\mathbf{x}-\mathbf{x}_a|$, we obtain long range exponential decay and Kato-like cusp condition at short range, similar to Slater orbitals. As a consequence, these basis functions could be particularly useful for all-electron calculations, where such a behavior is expected. This is indeed confirmed by results obtained in Section 3.3. However, in pseudopotential calculations, where the solution is expected to be smooth, $h_a(|\mathbf{x}-\mathbf{x}_a|) = |\mathbf{x}-\mathbf{x}_a|^2$ is a more appropriate choice. In this case, the resulting basis functions will have Gaussian type behavior, similar to Gaussian orbitals used in electronic structure calculations. It is also possible to vary $h_a(|\mathbf{x}-\mathbf{x}_a|)$ within the node set, which allows for the seamless transition between basis functions with different properties and behavior. For example, it is possible to easily transition between Gaussian-type basis functions and linear finite-elements within the domain (Arroyo and Ortiz (2006)).

In the previous chapter, we have proved the convergence of the finite-element approximation using the technique of $\Gamma-$ convergence. For this to be extended to the approximation scheme

presented here, the space generated by the basis functions needs to be dense in $H_0^1(\Omega)$. However, this result has so far only been proven for the choice $h_a(|\mathbf{x} - \mathbf{x}_a|) = |\mathbf{x} - \mathbf{x}_a|^2$ (Bompadre and Ortiz (2010)) and is still open for other choices of $h_a(|\mathbf{x} - \mathbf{x}_a|)$. Further, because of the nature of the basis functions, the convergence with numerical quadratures does not follow from previous work, and is still an open problem.

3.2 Numerical implementation

We now turn to the numerical implementation of the variational formulation described in Section 2.1. The variational problem (Eqn. (2.20)) is discretized using the *max-ent* basis functions introduced in Section 3.1

$$\psi_{i\sigma}(\mathbf{x}) = \sum_{j=1}^{P_h} \psi_{i\sigma}^j p_j(\mathbf{x}) \,, \tag{3.18}$$

$$\phi(\mathbf{x}) = \sum_{j=1}^{P_h} \phi^j p_j(\mathbf{x}) \tag{3.19}$$

to obtain

$$\sum_{j=1}^{P_h}\left[\int_\Omega\left(\frac{1}{2}\nabla p_j(\mathbf{x})\nabla p_k(\mathbf{x})+V^\sigma_{h,\text{eff}}(\mathbf{x},\mathbf{R})p_j(\mathbf{x})p_k(\mathbf{x})\right)\,\mathrm{d}\mathbf{x}\right]\psi^j_{i\sigma}$$
$$=\sum_{q=1}^{N_\sigma}\sum_{j=1}^{P_h}\lambda^q_{i\sigma}\psi^j_{q\sigma}\int_\Omega p_j(\mathbf{x})p_k(\mathbf{x})\,\mathrm{d}\mathbf{x}\,, \tag{3.20}$$

$$\sum_{j=1}^{P_h}\left[\frac{1}{4\pi}\int_\Omega\nabla p_k(\mathbf{x})\nabla p_j(\mathbf{x})\,\mathrm{d}\mathbf{x}\right]\phi^j=\int_\Omega(\rho_h(\mathbf{x})+b(\mathbf{x},\mathbf{R}))p_k(\mathbf{x})\,\mathrm{d}\mathbf{x}\,, \tag{3.21}$$

$$\sum_{s=1}^{P_h}\sum_{r=1}^{P_h}\left[\int_\Omega p_r(\mathbf{x})p_s(\mathbf{x})\,\mathrm{d}\mathbf{x}\right]\psi^r_{m\sigma}\psi^s_{n\sigma}=\delta_{mn} \tag{3.22}$$

where the Lagrange multipliers λ^σ_{iq} are used to enforce the constraints (Eqn. (2.21)). $V^\sigma_{h,\text{eff}}$, ρ_h denote the discretized V^σ_{eff}, ρ respectively and P_h represents the number of nodes in the node set. The equivalent generalized nonlinear eigenvalue problem corresponding to Eqns. (3.20) and (3.22) is given by

$$\mathbf{A}^\sigma\tilde{\Psi}^\sigma=\epsilon\mathbf{M}\tilde{\Psi}^\sigma,\qquad\sigma\in\{\alpha,\beta\}. \tag{3.23}$$

Here,

$$\mathbf{A}^\sigma{}_{ij} = \frac{1}{2}\int_\Omega\nabla p_i(\mathbf{x})\nabla p_j(\mathbf{x})\,\mathrm{d}\mathbf{x}+\int_\Omega V^\sigma_{\text{h,eff}}(\mathbf{x},\mathbf{R})p_i(\mathbf{x})p_j(\mathbf{x})\,\mathrm{d}\mathbf{x}\,, \tag{3.24}$$

$$\mathbf{M}_{ij} = \int_\Omega p_i(\mathbf{x})p_j(\mathbf{x})\,\mathrm{d}\mathbf{x}, \tag{3.25}$$

and $\tilde{\Psi}^\sigma$ is a vector of the nodal contribution to the wavefunction as given by Eqn. 3.1. The self-consistent field (SCF) method is commonly used for solving Eqns. (3.21) and (3.23).

The Newton-Raphson method is an attractive option for solving the nonlinear system of Eqns.

(3.20), (3.21) and (3.22) because of the quadratic rate of convergence achieved in the vicinity of the solution. However, the convergence rate is asymptotic and convergence is assured only if the initial guess lies within the radius of convergence of the solution. On the other hand, the SCF method is an expensive proposition on account of Eqn. (3.23) being a generalized eigenvalue problem.

Keeping the above discussion in mind, we employ the following methodology to solve the Kohn-Sham problem. We position the nodes at a geometric coarsening rate away from the nuclei to enhance the efficiency of the calculations. We enforce zero Dirichlet boundary conditions for both the wavefunctions and electrostatic potential by utilizing the Kronecker-delta property satisfied by the basis functions on the boundary of Ω. We incorporate the best attributes of the SCF and Newton-Raphson methods into our solution procedure along the lines of our previous work (Suryanarayana et al. (2010)). We rewrite Eqn. (3.23) as

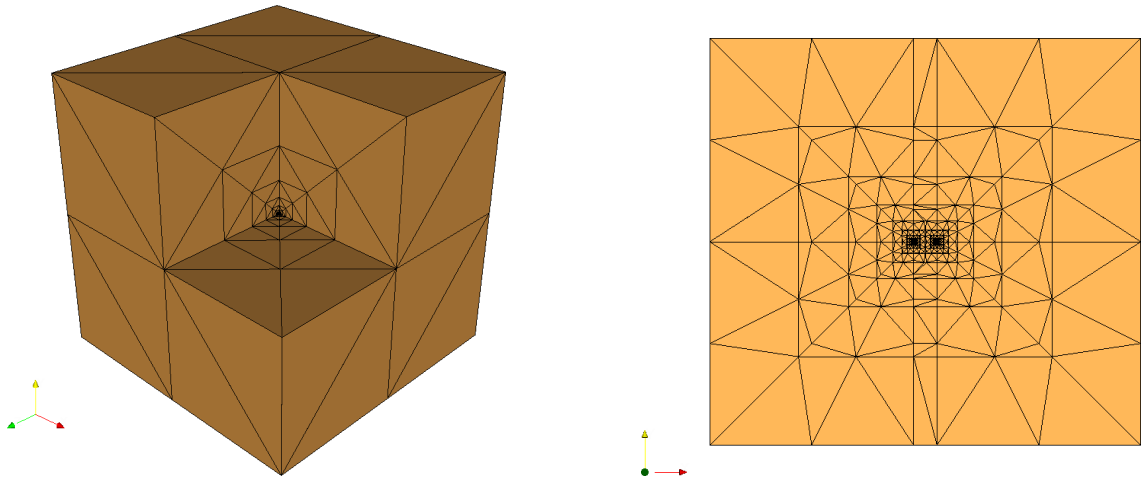
$$(\mathbf{A}^{\sigma} - \eta\mathbf{M})^{-1}\mathbf{M}\tilde{\Psi}^{\sigma} = \hat{\epsilon}\tilde{\Psi}^{\sigma} \tag{3.26}$$

where η is the shift parameter and $\hat{\epsilon} = \frac{1}{\epsilon-\eta}$, thereby magnifying the region of interest in the eigenvalue spectrum resulting in significant increase in the rate of convergence without incurring any additional cost. We use a low resolution node set Y_0 to solve Eqns. (3.21) and (3.26) using the SCF method. Specifically, we employ the implicitly restarted Lanczos method (Saad (1992)) for solving the linear eigenvalue problem and the conjugate gradient algorithm as the linear solver. Since Y_0 has low resolution, the computational expense involved for this step is minimal. Next, we generate Y_1 which is a refinement of the node set Y_0. The solutions obtained on Y_0 are transferred to Y_1 using the basis functions of Y_0. This serves as a starting guess for the solution of Eqns. (3.20), (3.21)

and (3.22) simultaneously using Newton's method with Goldstein-Armijo line-searches (Dennis and Schnabel (1996)). For the associated linear solver we use the generalized minimal residual method (GMRES, Saad and Schultz (1986)), while retaining the option of switching to either the bi-conjugate gradient stable method (Bi-CGSTAB, van der Vorst (1992)) or transpose-free quasi-minimal residual method (TFQMR, Freund (1993)). Since the Hessian information is required only through its product with a vector, we evaluate it by using directional derivative quotients (Brown (1987); Brown and Saad (1990)) which is significantly faster than evaluating the exact Hessian. The quality of the initial guess generated by the procedure just described allows us to obtain convergence within exceedingly tight tolerances in no more than 6 to 7 iterations at a quadratic rate.

The equilibrium position of the nuclei are computed using Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. Each force update following a displacement of the nuclei requires the re-evaluation of the electrostatic potential and wavefunctions within an internal loop. In order to expedite this re-evaluation, we use the solution from the previous configuration as an initial guess for the Newton-Raphson method. In order to verify convergence with respect to number of basis functions, we repeat the calculations on increasingly finer node sets Y_n and the solution on the parent node set is used as starting guess for the Newton-Raphson method on the finer node set, which ensures rapid convergence. This process is repeated until convergence is achieved to within a prespecified tolerance. The recursive nature of the calculations is advantageous with respect to solution schemes that require a complete restart every time the discretization is modified, e.g., in plane wave basis calculations, the convergence has to be verified with respect to the energy cut-off and number of k-points sampled.

We evaluate all the integrals via Gaussian quadrature (Solin and Dolezel (2004)) on the Delaunay triangulation of Ω generated using the corresponding *max-ent* node set Y_n. The coarsening nature of the triangulation allows for efficient and accurate evaluation of the integrals. See Fig. 3.1 for examples of triangulations used for a single atom and dimer. At each quadrature point, we solve the minimization problem (Eqn. (3.12)) using the Newton-Raphson method to obtain values of the shape functions and calculate their derivatives using the expressions derived in Appendix B. It should be noted that since an exact quadrature rule for the *max-ent* basis functions does not exist, we verify the convergence of the solution with respect to the order of the quadrature used. In order to reduce the error due to the inexact quadrature rule, we use the same node sets for evaluating the energy of both the cluster of atoms as well as the individual component atoms. This results in significant quadrature error cancellation.



(a) Single atom.

(b) Dimer.

Figure 3.1: Examples of sliced background triangulations used for numerical integration

Finally, the implementation is parallelized as follows. We decompose the cubical domain Ω into

cubical subdomains Ω^i and correspondingly the node set Y into subsets Y^i. In order to satisfy the *conforming patches* property (Arroyo and Ortiz (2006)), the node sets Y, Y^i should be chosen such that Ω^i is the convex hull of Y^i. Each node set Y^i and corresponding subdomain Ω^i is now handled independently by a different processor. Since the basis functions for Y^i satisfy the Kronecker-delta property on the boundary of Ω^i (Arroyo and Ortiz (2006)), inter-processor communication is required only for the nodes on the boundary. This is exactly the same situation when performing parallel computations using finite-elements via domain decomposition. Therefore, efficient parallel computations can be performed, akin to simplicial finite-elements.

3.3 Examples and results

In this section, we validate our implementation through a number of examples ranging from single atoms to small clusters of atoms. We do so for both the all-electron case as well as local and nonlocal pseudopotential approximations. For all-electron calculations, we choose $h_a(|\mathbf{x}-\mathbf{x}_a|) = |\mathbf{x}-\mathbf{x}_a|$, which results in basis functions with the expected behavior of the solution inherent to them. For the pseudopotential calculations, we verify the obtained results with both $h_a(|\mathbf{x}-\mathbf{x}_a|) = |\mathbf{x}-\mathbf{x}_a|$ and $|\mathbf{x}-\mathbf{x}_a|^2$. We vary $\beta = \gamma/h, \gamma/h^2$ for the choice of $h_a(|\mathbf{x}-\mathbf{x}_a|) = |\mathbf{x}-\mathbf{x}_a|, |\mathbf{x}-\mathbf{x}_a|^2$ respectively, where γ is a dimensionless constant and h is the representative spacing between the nodes.

3.3.1 All-electron calculations

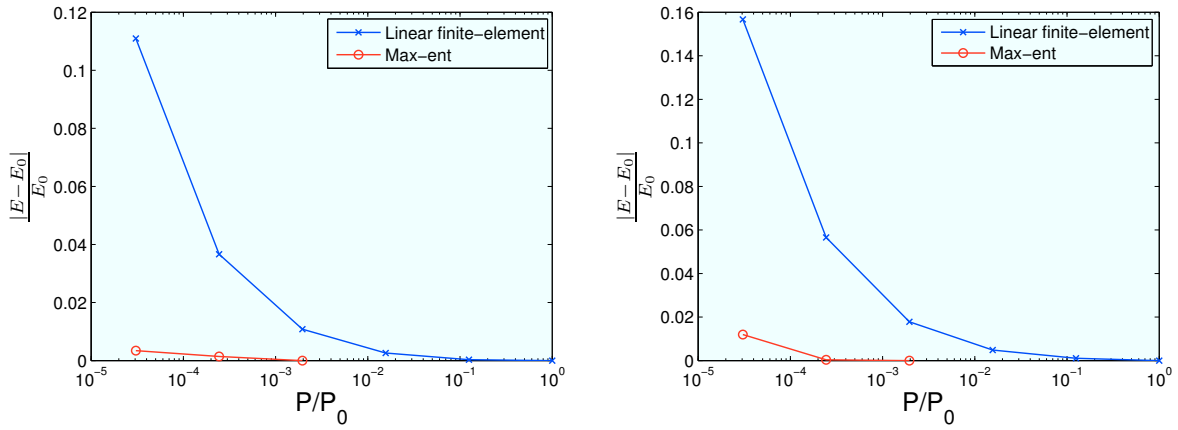
First we perform all-electron single atom calculations for hydrogen, helium, lithium, beryllium and carbon. The results obtained have been compared with highly accurate calculations of Kotochigova et al. (1997) in Table 3.1. We obtain very good agreement validating the accuracy of our implementation.

Table 3.1: Ground state energy of selected atoms (a.u)

Element	Max-ent	Kotochigova et al.
H	-0.445	-0.445
He	-2.830	-2.834
Li	-7.338	-7.335
Be	-14.434	-14.447
C	-37.432	-37.425

Figs. 3.2a and 3.2b provide a comparison of the convergence obtained using *max-ent* basis functions with simplicial linear finite-elements (Suryanarayana et al. (2010)). Note that the position of the nodes correspond to the optimal triangulations developed for the simplicial linear finite-elements (Suryanarayana et al. (2010)). E_0 denotes the converged energy and P_0 denotes the number of nodes required for the converged finite-element solution. For the helium, carbon atoms we have used 1250, 10000 *max-ent* basis functions to achieve a converged solution. It is clear that there is a significant reduction in the number of basis functions required to achieve convergence, resulting in computational times differing by an order of magnitude, inspite of their non-locality and necessity for higher order quadrature rules. The relatively small number of basis functions make them a viable choice for performing medium sized all-electron calculations.

The next set of examples considered are the nitrogen (N_2) and carbon monoxide (CO) molecules.



(a) Helium atom.

(b) Carbon atom.

Figure 3.2: Comparison between the convex approximation scheme and simplicial linear finite-elements for all-electron calculations

A comparison of the obtained binding energy and bond length with previous studies is provided in Tables 3.2 and 3.3. Again, the agreement is good.

Table 3.2: Binding energy and bond length of N_2

Property	Max-ent	Engel et al. (2001)	Suryanarayana et al. (2010)
Binding energy (eV)	-11.6	-11.593	-11.6
Bond length (a.u.)	2.06	2.068	2.06

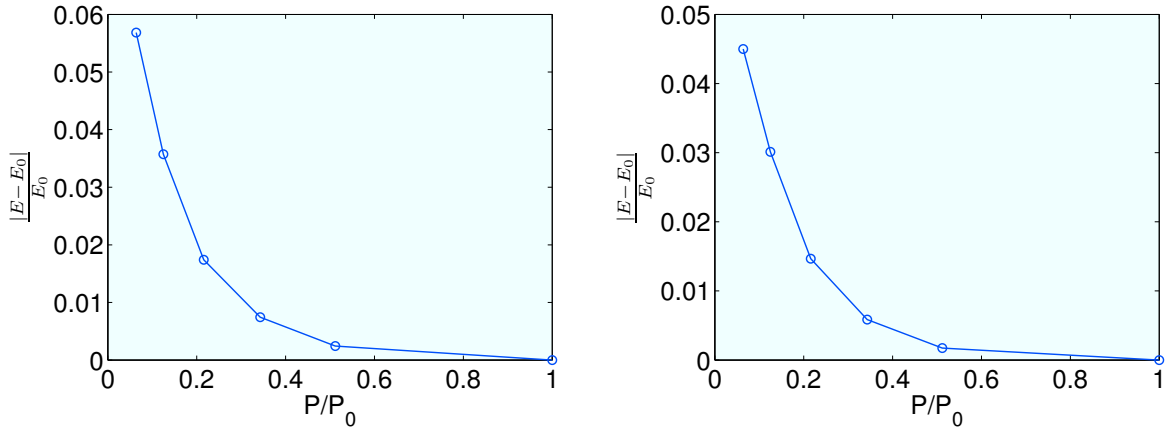
Table 3.3: Binding energy and bond length of CO

Property	Max-ent	Engel et al. (2001)	Suryanarayana et al. (2010)
Binding energy (eV)	-13.0	-12.967	-13.03
Bond length (a.u.)	2.10	2.128	2.08

3.3.2 Local pseudopotential approximation

In this section we present the results obtained using the local 'Evanescent Core' (EC) pseudopotential (Fiolhais et al. (1995)). First, we provide plots of convergence of the energy for the sodium

atom using $h_a(|\mathbf{x}-\mathbf{x}_a|) = |\mathbf{x}-\mathbf{x}_a|, |\mathbf{x}-\mathbf{x}_a|^2$ in Figs 3.3a, 3.3b respectively. For this convergence study, we have used uniformly spaced nodes with $P_0 = 8000$. Next, we consider dimers of lithium and sodium, the results of which are presented in Table 3.4. The agreement is good and the difference from Nogueira et al. (1996) arises because of their use of the linear combination of atomic orbitals (LCAO) theory. Electron density contours on the mid-plane of Na_2 has been plotted in Fig. 3.4a.



(a) $h_a(|\mathbf{x}-\mathbf{x}_a|) = |\mathbf{x}-\mathbf{x}_a|$.

(b) $h_a(|\mathbf{x}-\mathbf{x}_a|) = |\mathbf{x}-\mathbf{x}_a|^2$.

Figure 3.3: Convergence of energy of the sodium atom using the EC pseudopotential for different choices of $h_a(|\mathbf{x}-\mathbf{x}_a|)$

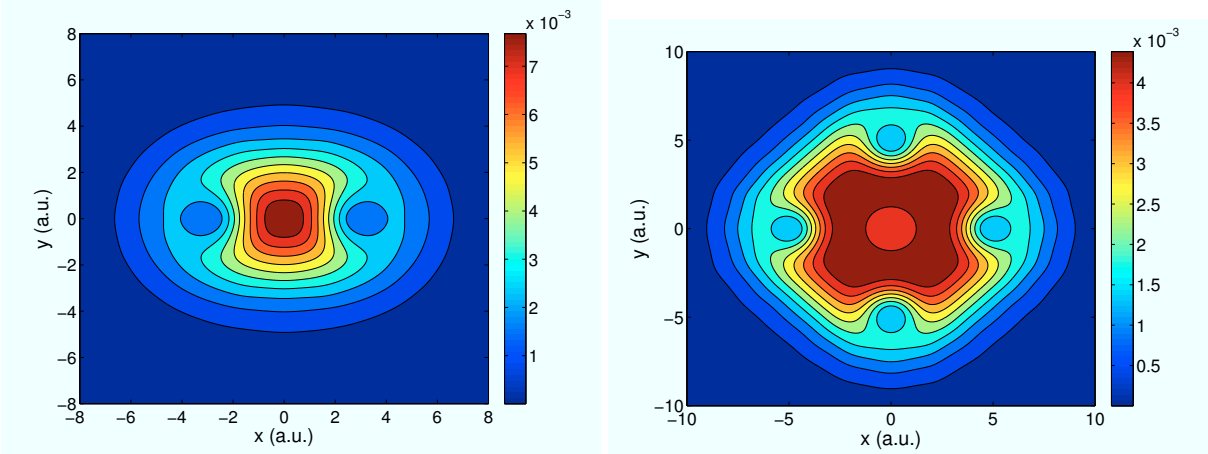
Next, we study octahedral clusters of lithium and sodium. Note that we do not try to optimize the geometry with respect to different arrangements of the atoms and we neglect spin polarization of the cluster to facilitate comparison with previous studies. Table 3.5 lists the binding energy and bond length so obtained. Again, the difference in results obtained by Nogueira et al. (1996) arises because of their use of LCAO theory. We plot electron density contours on the mid-plane of Na_6 in Fig. 3.4b.

Table 3.4: Binding energy and bond length for dimers of lithium and sodium using the EC pseudopotential

Dimer	Property	Max-ent	Suryanarayana et al. (2010)	Nogueira et al. (1996)
Li_2	Binding energy (eV/atom)	-0.49	-0.49	-0.52
	Bond length (a.u)	4.82	4.86	4.92
Na_2	Binding energy (eV/atom)	-0.36	-0.35	-0.46
	Bond length (a.u)	5.72	5.72	5.77

Table 3.5: Binding energy and bond length for octahedral clusters of lithium and sodium using the EC pseudopotential

Dimer	Property	Max-ent	Nogueira et al. (1996)	Fiolhais et al. (1996)
Li_6	Binding energy (eV/atom)	-0.5	-0.72	-
	Bond length (a.u)	5.66	5.79	-
Na_6	Binding energy/atom (eV/atom)	-0.42	-0.53	-0.67
	Bond length (a.u)	6.78	6.87	6.31



(a) Na_2 (b) Na_6

Figure 3.4: Contours of electron density on the mid-plane of sodium clusters using the EC pseudopotential

Finally, we study a cluster of $3 \times 3 \times 3$ body centered cubic (BCC) unit cells of lithium with

fixed geometry. The binding energy and lattice constant obtained are listed in Table 3.6. Contours of electron density on the mid-plane of this cluster of atoms can be seen in Fig. 3.5.

Table 3.6: Binding energy and lattice constant of $3 \times 3 \times 3$ BCC unit cells of lithium using the EC pseudopotential

Property	Max-ent
Binding energy (eV/atom)	-0.97
Lattice constant (a.u.)	6.40

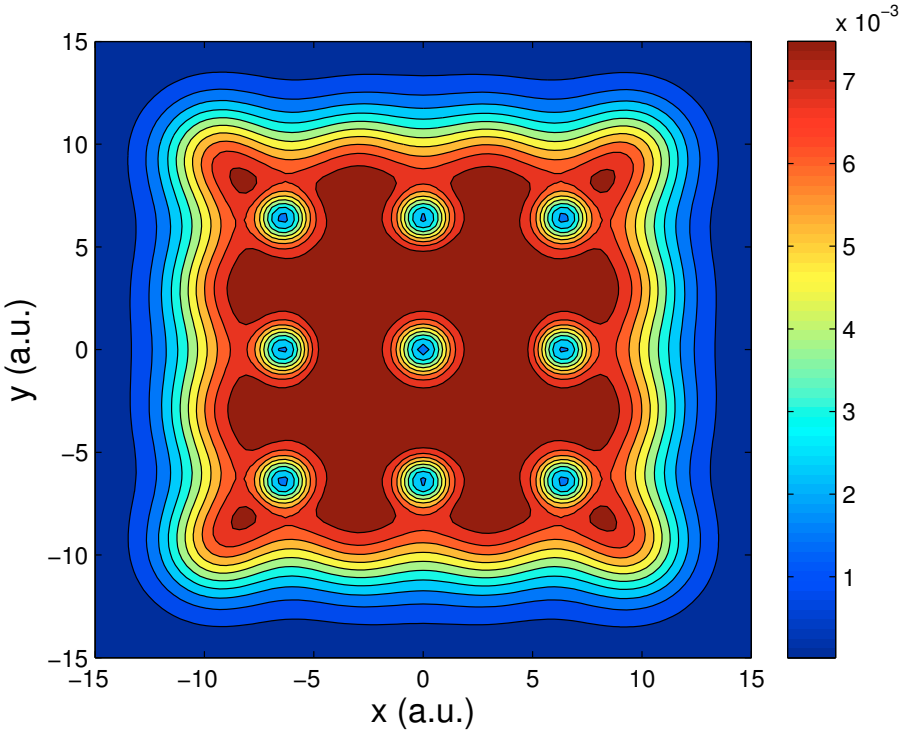


Figure 3.5: Contours of electron density on the mid-plane of $3 \times 3 \times 3$ BCC unit cells of lithium obtained using the EC pseudopotential

Chapter 4

Linear-scaling spectral Gauss quadrature method

Traditional implementations of DFT solve for the wavefunctions (Chelikowsky et al. (1994); Kresse and Furthmüller (1996); Pask et al. (1999); Ismail-Beigi and Arias (2000); Segall et al. (2002); Gonze et al. (2002); Tsuchida (2004); Castro et al. (2006); Suryanarayana et al. (2010, 2011)), a procedure which has cubic-scaling with respect to the number of atoms. This restricts the size of the systems which can be studied to a few hundred atoms. To overcome this, there have been significant efforts towards developing linear-scaling methods for DFT (García-Cervera et al. (2007); Garcia-Cervera et al. (2009); Galli and Parrinello (1992); Mauri et al. (1993); Skylaris et al. (2005); Barrault et al. (2007)). For a more complete survey on the topic, please see Goedecker (1999). However, these methods are found to be linear-scaling only for insulators and developing a linear-scaling method for metals is a challenging open problem (Cances et al. (2008)). In this chapter, we develop a linear-scaling method for DFT which is nominally linear-scaling for both insulating and metallic systems. Further, this method lays the groundwork for coarse-graining which we later discuss in Chapter 5.

The outline of the rest of the chapter is as follows. In Section 4.1, we reformulate DFT in terms

of the electron density and the band structure energy. We provide some mathematical background on spectral theory and Gauss quadrature in Section 4.2. Next, we describe the proposed method in Section 4.3 which we validate the formulation through examples in Section 4.4. We also discuss the connection of the proposed method with the so called recursion method (Haydock (1980)) and Padé approximation in Appendix C.

4.1 Density Functional Theory - Eigenvalue problem

The Kohn-Sham problem (Kohn and Sham (1965)) can be expressed as a nonlinear eigenvalue problem (cf. e.g. Parr and Yang (1989))

$$\mathcal{H}\psi_n = \lambda_n \psi_n \,, \quad n = 1, 2, \ldots \tag{4.1}$$

where

$$\mathcal{H} = -\frac{1}{2}\nabla^2 + V_{ext}(\mathbf{x}, \mathbf{R}) + V_H[\rho] + V_{xc}[\rho]$$

is a self-adjoint operator with eigenvalues λ_n and corresponding orthonormal eigenfunctions ψ_n. The electron density can in turn be expressed in terms of the eigenfunctions as

$$\rho(\mathbf{x}) = 2\sum_n g(\lambda_n, \lambda_f)|\psi_n(\mathbf{x})|^2 \tag{4.2}$$

where

$$g(\lambda, \lambda_f) = \begin{cases} 1, & \text{if } \lambda \leq \lambda_f \\ 0, & \text{otherwise} \end{cases} \tag{4.3}$$

is the orbital occupation function and λ_f is the Fermi energy. Note that the factor of two arises because of the two spins the electron can possess. The Fermi energy is evaluated by solving for the constraint

$$2\sum_n g(\lambda_n, \lambda_f) = N_e \tag{4.4}$$

where N_e is the total number of electrons.

The external potential due to the nuclei is given by

$$V_{ext}(\mathbf{x}, \mathbf{R}) = \sum_{J=1}^{M} \frac{Z_J}{|\mathbf{x} - \mathbf{R}_J|} \tag{4.5}$$

where $\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2, \ldots \mathbf{R}_M\}$ are the positions of the nuclei and $Z = \{Z_1, Z_2, \ldots Z_M\}$ are the corresponding charges of the nuclei. Due to the chemically intert nature of the core electrons and the large computational expense involved in capturing the oscillations of the valence wavefunctions in the core region, it is common to replace the singular Coloumb potential in Eqn. (4.5) by an effective potential

$$V_{ext}^{PS}(\mathbf{x}, \mathbf{R})\psi_n(\mathbf{x}) = V_{nl}(\mathbf{x}, \mathbf{R})\psi_n(\mathbf{x}) + V_{loc}(\mathbf{x}, \mathbf{R})\psi_n(\mathbf{x}) \tag{4.6}$$

where $V_{nl}(\mathbf{x}, \mathbf{R})$ and $V_{loc}(\mathbf{x}, \mathbf{R})$ are the nonlocal and local parts of the pseudopotential, respectively. Usually, $V_{nl}(\mathbf{x}, \mathbf{R})$ is non-zero only in a small region around each nucleus and $V_{loc}(\mathbf{x}, \mathbf{R})$

replicates the Coloumb potential outside these core regions. In this approximation scheme, the core electrons are no longer considered as their effect is modeled by $V_{ext}^{PS}(\mathbf{x}, \mathbf{R})$, which also serves the dual purpose of ensuring nodeless valence pseudo-wavefuntions. This is commonly recognized as the pseudopotential approximation (cf. e.g. Martin (2004)). Highly successful pseudopotentials include the norm conserving (Bachelet et al. (1982); Rappe et al. (1990); Troullier and Martins (1991)) and ultrasoft pseudopotentials (Vanderbilt (1990)).

The electrostatic potential arising from the electron density, also referred to as the Hartree potential, can be represented as

$$V_H[\rho](\mathbf{x}) = \int_{\mathbb{R}^3} \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \, \mathrm{d}\mathbf{x}'. \tag{4.7}$$

The exchange-correlation potential is defined to be the first variation of the exchange-correlation energy $E_{xc}[\rho]$

$$V_{xc}[\rho] = \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{x})}. \tag{4.8}$$

Commonly used models for the exchange-correlation energy include the local density approximation (LDA) (Kohn and Sham (1965)) and generalized gradient approximation (GGA) (Langreth and Mehl (1983); Perdew et al. (1992)).

The potential due to the nuclei ($V_{ext}(\mathbf{x}, \mathbf{R})$ or $V_{loc}(\mathbf{x}, \mathbf{R})$) and the Hartree potential can be cast into a local form as the solution to the Poisson equation (Ismail-Beigi and Arias (2000); Suryanarayana et al. (2010))

$$-\frac{1}{4\pi}\nabla^2\phi(\mathbf{x}, \mathbf{R}) = \rho(\mathbf{x}) + b(\mathbf{x}, \mathbf{R}) \tag{4.9}$$

which has the unique solution

$$\phi(\mathbf{x},\mathbf{R}) = \int_{\mathbb{R}^3} \frac{\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}\,\mathrm{d}\mathbf{x}' + \int_{\mathbb{R}^3} \frac{b(\mathbf{x}',\mathbf{R})}{|\mathbf{x}-\mathbf{x}'|}\,\mathrm{d}\mathbf{x}'. \tag{4.10}$$

Here, $b(\mathbf{x}',\mathbf{R}) = \sum_{J=1}^{M} b_J(\mathbf{x}',\mathbf{R}_J)$ denotes the total charge distribution of the nuclei, with $b_J(\mathbf{x}',\mathbf{R}_J)$ representing the charge density of the J^{th} nucleus.

Thereafter, for fixed position of the nuclei, the Kohn-Sham problem takes the form

$$\begin{aligned}
\mathcal{H}\psi_n &= \lambda_n\psi_n\,, \quad \mathcal{H} = -\frac{1}{2}\nabla^2 + V_{xc}[\rho] + \phi(\mathbf{x},\mathbf{R}) + V_{nl}(\mathbf{x},\mathbf{R}) && (4.11)\\
-\frac{1}{4\pi}\nabla^2\phi(\mathbf{x},\mathbf{R}) &= \rho(\mathbf{x}) + b(\mathbf{x},\mathbf{R}) && (4.12)\\
\rho(\mathbf{x}) &= 2\sum_n g(\lambda_n,\lambda_f)|\psi_n(\mathbf{x})|^2 && (4.13)\\
2\sum_n g(\lambda_n,\lambda_f) &= N_e && (4.14)
\end{aligned}$$

which can be self-consistently using the self consistent field (SCF) method (cf. e.g. Martin (2004)). Note that $V_{nl}(\mathbf{x},\mathbf{R}) \equiv 0$ for an all-electron calculation and on the choice of a fully local pseudopotential. On solving these equations, we can evaluate the energy of the system for the given position of the nuclei

$$\begin{aligned}
\mathcal{E}_0 &= 2\sum_n g(\lambda_n,\lambda_f)\lambda_n + \frac{1}{2}\int_{\mathbb{R}^3}(b(\mathbf{x},\mathbf{R}) - \rho(\mathbf{x}))\phi(\mathbf{x},\mathbf{R})\,\mathrm{d}\mathbf{x} + \mathrm{E_{xc}}[\rho]\\
&\quad - \int_{\mathbb{R}^3} V_{xc}[\rho]\rho(\mathbf{x})\,\mathrm{d}\mathbf{x} - \frac{1}{2}\sum_{\mathrm{J}=1}^{\mathrm{M}}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\frac{\mathrm{b_J}(\mathbf{x},\mathbf{R}_\mathrm{J})\mathrm{b_J}(\mathbf{x}',\mathbf{R}_\mathrm{J})}{|\mathbf{x}-\mathbf{x}'|}\,\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}' \qquad (4.15)
\end{aligned}$$

where the first term represents the band structure energy (U_{band}) and the last term is the self energy

of the nuclei. To calculate the ground state energy of the system, we need to further minimize the energy with respect to the positions of the nuclei which can be accomplished by equilibriating the force on the nuclei

$$f_J = -\frac{\partial \mathcal{E}_0}{\partial \mathbf{R}_J}, \quad J = 1, 2 \dots M. \tag{4.16}$$

For a fixed position of the nuclei, if the energy is stationary with respect to the electron density and electrostatic potential, it follows that

$$f_J = -\int_{\mathbb{R}^3} \frac{db_J(\mathbf{x}, \mathbf{R}_J)}{d\mathbf{R}_J} \left(\phi(\mathbf{x}, \mathbf{R}) - \phi_J(\mathbf{x}, \mathbf{R}_J)\right) \, \mathrm{d}\mathbf{x} \tag{4.17}$$

where $\phi_J(\mathbf{x}, \mathbf{R}_J) = \int_{\mathbb{R}^3} \frac{b_J(\mathbf{x}', \mathbf{R}_J)}{|\mathbf{x} - \mathbf{x}'|} \, \mathrm{d}\mathbf{x}'$. In the case of all-electron calculations where we have $b_J(\mathbf{x}, \mathbf{R}_J) = -Z_J \delta(\mathbf{x} - \mathbf{R}_J)$, we obtain

$$f_J = Z_J \nabla(\phi(\mathbf{x}, \mathbf{R}) - \phi_J(\mathbf{x}, \mathbf{R}_J))\Big|_{\mathbf{x}=\mathbf{R}_J} \tag{4.18}$$

This result is commonly referred to as the Hellmann-Feynman theorem (cf. e.g. Finnis (2003)).

In the above discussion of the Kohn-Sham method, we have inherently assumed a temperature of absolute zero. It can be extended to finite temperatures (cf. e.g. Parr and Yang (1989)) by introducing the Helmholtz free energy

$$\mathcal{F} = \mathcal{E}_\sigma - \theta \mathcal{S} \tag{4.19}$$

where $\mathcal{E}_\sigma$ is the finite temperature counterpart of the ground state energy, θ is the temperature and

$\mathcal{S}$ is the entropy. The orbital occupation function can now take fractional values as defined by the Fermi-Dirac distribution

$$g(\lambda, \lambda_f) = \frac{1}{1 + \exp(\frac{\lambda - \lambda_f}{\sigma})} \tag{4.20}$$

where $\sigma = k_B\theta$, k_B being the Boltzmann constant. We also have the following representation for the entropy

$$\mathcal{S} = -2k_B \sum_n \left[g(\lambda_n, \lambda_f) \log g(\lambda_n, \lambda_f) + (1 - g(\lambda_n, \lambda_f) \log(1 - g(\lambda_n, \lambda_f))) \right]. \tag{4.21}$$

The forces on the nuclei

$$f_J = -\frac{\partial \mathcal{F}}{\partial \mathbf{R}_J}. \tag{4.22}$$

reduces to the expression given by Eqn. (4.17) (Weinert and Davenport (1992)). Using the finite temperature calculation, it is also possible to to get an accurate extrapolation for the ground state energy at absolute zero (Gillan (1989))

$$\mathcal{E}_0 \approx \frac{1}{2} \left(\mathcal{E}_\sigma + \mathcal{F} \right). \tag{4.23}$$

4.2 Mathematical background

4.2.1 Spectral theory

Let $\mathcal{B}(H)$ denote the Banach algebra of all self-adjoint linear operators $\mathcal{H}$ on a finite-dimensional Hilbert space H with inner product $(.,.)$ and norm $\|.\|$. We have for $\mathcal{H} \in \mathcal{B}(H)$

$$\|\mathcal{H}\| = \sup\{\|\mathcal{H}\eta\| : \eta \in H,\ \|\eta\| \leq 1\}. \tag{4.24}$$

In this work $H = L^2(\Omega)$, the space of square integrable functions over Ω. From the spectral theorem (cf. e.g. Rudin (1991)), it follows that we have a unique resolution of the identity $\mathcal{E}$ on the Borel subsets of its spectrum $\sigma(\mathcal{H}) = \{\lambda_1, \lambda_2, \ldots, \lambda_{N_d}\} \subset \mathbb{R}$ which satisfies

$$\mathcal{H} = \int_{\sigma(\mathcal{H})} \lambda \, d\mathcal{E}(\lambda). \tag{4.25}$$

Consequently for any bounded Borel function f on $\sigma(\mathcal{H})$, we have the following representation

$$f(\mathcal{H}) = \int_{\sigma(\mathcal{H})} f(\lambda) \, d\mathcal{E}(\lambda). \tag{4.26}$$

Let the eigenvalues of $\mathcal{H}$ be ordered according to $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{N_d}$. Further, let ψ_n denote the eigenfunction corresponding to the eigenvalue λ_n. We have the following representation of the

Borel measure on $\sigma(\mathcal{H})$

$$\mathcal{E}_{\zeta,\zeta}(\lambda) = (\mathcal{E}(\lambda)\zeta,\zeta) = \begin{cases} 0, & \text{if } \lambda < \lambda_1 \\ \sum_{n=1}^{m}\sum_{p=1}^{N_d}\sum_{q=1}^{N_d}\psi_{n,p}\psi_{n,q}\zeta_p\zeta_q, & \text{if } \lambda_m \le \lambda < \lambda_{m+1} \\ \sum_{n=1}^{N_d}\sum_{p=1}^{N_d}\sum_{q=1}^{N_d}\psi_{n,p}\psi_{n,q}\zeta_p\zeta_q, & \text{if } \lambda_{N_d} < \lambda \end{cases} \tag{4.27}$$

where $\zeta = \sum_{p=1}^{N_d}\zeta_p\eta_p$ and $\psi_n = \sum_{p=1}^{N_d}\psi_{n,p}\eta_p$. Here, $\{\eta_k\}_{k=1}^{N_d}$ is an orthonormal basis for H. In the special case of $\zeta = \eta_k$, it reduces to

$$\mathcal{E}_{\eta_k,\eta_k}(\lambda) = \begin{cases} 0, & \text{if } \lambda < \lambda_1 \\ \sum_{n=1}^{m}|\psi_{n,k}|^2, & \text{if } \lambda_m \le \lambda < \lambda_{m+1} \\ \sum_{n=1}^{N_d}|\psi_{n,k}|^2, & \text{if } \lambda_{N_d} < \lambda. \end{cases} \tag{4.28}$$

Therefore $\forall\zeta \in H$

$$(f(\mathcal{H})\zeta,\zeta) = \int_{\sigma(\mathcal{H})} f(\lambda)\, d\mathcal{E}_{\zeta,\zeta}(\lambda) = \int_a^b f(\lambda)\, d\mathcal{E}_{\zeta,\zeta}(\lambda) \tag{4.29}$$

for $a = \lambda_1$ and $b = \lambda_{N_d}$.

4.2.2 Gauss quadrature

Let $\mathcal{P}$ be the space of real polynomials and $\mathcal{P}_k$ be a subspace of $\mathcal{P}$ consisting of polynomials of degree at most k. We define an inner product (relative to the measure $\mathcal{E}_{\zeta,\zeta}$) of two polynomials

$p, q \in \mathcal{P}$ as

$$\langle p, q \rangle_\zeta = \int_a^b p(\lambda) q(\lambda)\, d\mathcal{E}_{\zeta,\zeta}(\lambda) \tag{4.30}$$

with norm

$$\|p\|_\zeta = \left(\int_a^b p^2(\lambda)\, d\mathcal{E}_{\zeta,\zeta}(\lambda) \right)^{\frac{1}{2}}. \tag{4.31}$$

Here, we are interested in approximating integrals of the form

$$I[f] = \int_a^b f(\lambda)\, d\mathcal{E}_{\zeta,\zeta}(\lambda) \tag{4.32}$$

using a quadrature rule. To do so, we approximate the function $f(\lambda)$ with an interpolation polynomial

$$f(\lambda) \approx \sum_{k=1}^{K} f(\lambda_k^\zeta) l_k^\zeta(\lambda) \tag{4.33}$$

where $\{\lambda_k^\zeta\}_{k=1}^K$ are the nodes/interpolation points and $l_k^\zeta(\lambda)$ is the Lagrange polynomial

$$l_k^\zeta(\lambda) = \prod_{j=1, j\neq k} \frac{\lambda - \lambda_j^\zeta}{\lambda_k^\zeta - \lambda_j^\zeta}. \tag{4.34}$$

Thereafter, we obtain the quadrature formula

$$I[f] \approx \sum_{k=1}^{K} w_k^\zeta f(\lambda_k^\zeta) \tag{4.35}$$

where

$$w_k^\zeta = \int_a^b l_k^\zeta(\lambda) d\mathcal{E}_{\zeta,\zeta}(\lambda) \tag{4.36}$$

are the weights of the quadrature rule. In order to obtain the quadrature rule of highest degree, we further consider the nodes $\{\lambda_k\}_{k=1}^K$ as unknowns. This results in a quadrature rule of degree $2K-1$ i.e. any $p \in \mathcal{P}_{2K-1}$ is integrated exactly, and is referred to as Gauss quadrature. To efficiently evaluate the nodes and weights of the Gauss quadrature rule, we employ the procedure described below (cf. e.g. Golub and Meurant (2010)).

We can generate a sequence of orthonormal polynomials (with respect to the measure $\mathcal{E}_{\zeta,\zeta}$) $\{\hat{p}_k\}_{k=0}^K$ through a three term recurrence relationship

$$\begin{aligned} b_{k+1}\hat{p}_{k+1}(\lambda) = (\lambda - a_{k+1})\hat{p}_k(\lambda) - b_k\hat{p}_{k-1}(\lambda)\,, \quad k = 0, 1, \ldots, K-1 \\ \hat{p}_{-1}(\lambda) = 0\,, \quad \hat{p}_0(\lambda) = 1\,, \quad b_0 = 1 \end{aligned} \tag{4.37}$$

where

$$a_{k+1} = \langle \lambda\hat{p}_k, \hat{p}_k \rangle_\zeta\,, \quad k = 0, 1, \ldots, K-1 \tag{4.38}$$

and $\hat{b}_k$ is computed such that $\|\hat{p}_k\|_\zeta = 1\,,\ k = 0, 1, \ldots, K$. Corresponding to these orthonormal polynomials, there is a tridiagonal Jacobi matrix $\hat{J}_K$ of dimension K

$$\hat{J}_K = \begin{pmatrix} a_1 & b_1 & & & \\ b_1 & a_2 & b_2 & & \\ & \ddots & \ddots & \ddots & \\ & & b_{K-2} & a_{K-1} & b_{K-1} \\ & & & b_{K-1} & a_K \end{pmatrix}. \tag{4.39}$$

Let us denote $\hat{P}_K(\lambda) = (\hat{p}_0(\lambda), \hat{p}_1(\lambda), \ldots, \hat{p}_{K-1}(\lambda))^T$. Then the three term recurrence relation given by Eqn. (4.37) can be written compactly as

$$\lambda \hat{P}_K(\lambda) = \hat{J}_K(\lambda)\hat{P}_K(\lambda) + b_K \hat{p}_K(\lambda) e_K \tag{4.40}$$

where e_K is the last column of the identity matrix of dimension K. It can be shown that the eigenvalues of $\hat{J}_K$ (which are also the zeros of $\hat{p}_K(\lambda)$) are the nodes $\{\lambda_j\}_{j=1}^K$ of the Gauss quadrature rule and the weights $\{w_j\}_{j=1}^K$ are the squares of the first elements of the normalized eigenvectors.

4.3 Formulation

4.3.1 Integral representations

We rewrite the expressions for the Fermi energy, electron density, band structure energy and entropy as Riemann-Stieltjes integrals over the spectrum of $\mathcal{H} \in \mathcal{B}(H)$. This enables the use of Gauss quadrature for their evaluation. This is a key ingredient in the formulation of LSSGQ, further described in Section 4.3.2.

4.3.1.1 Fermi energy

The Fermi energy is calculated by solving for the constraint

$$2\sum_{n=1}^{N_d} g(\lambda_n, \lambda_f) = N_e. \tag{4.41}$$

Since $\|\psi_n\| = \sum_{p=1}^{N_d} |\psi_{n,p}|^2 = 1$, it follows that

$$\begin{aligned}\sum_{n=1}^{N_d} g(\lambda_n, \lambda_f) &= \sum_{p=1}^{N_d}\sum_{n=1}^{N_d} g(\lambda_n, \lambda_f)|\psi_{n,p}|^2 \\ &= \sum_{p=1}^{N_d}\int_a^b g(\lambda, \lambda_f) d\mathcal{E}_{\eta_p,\eta_p}(\lambda). \end{aligned} \tag{4.42}$$

Therefore, Eqn. (4.41) can be rewritten as

$$2\sum_{p=1}^{N_d}\int_a^b g(\lambda, \lambda_f) d\mathcal{E}_{\eta_p,\eta_p}(\lambda) = N_e. \tag{4.43}$$

4.3.1.2 Electron density

The electron density at any point $\mathbf{x}_0$

$$\begin{aligned}\rho(\mathbf{x}_0) &= 2\sum_{n=1}^{N_d} g(\lambda_n, \lambda_f)|\psi_n(\mathbf{x}_0)|^2 \\ &= 2\sum_{n=1}^{N_d}\sum_{p=1}^{N_d}\sum_{q=1}^{N_d} g(\lambda_n, \lambda_f)\psi_{n,p}\psi_{n,q}\eta_p(\mathbf{x}_0)\eta_q(\mathbf{x}_0) \\ &= 2\int_a^b g(\lambda, \lambda_f)\, d\mathcal{E}_{\zeta,\zeta}(\lambda) \end{aligned} \tag{4.44}$$

where $\zeta = \sum_p \eta_p(\mathbf{x}_0)\eta_p$.

4.3.1.3 Band structure energy

The band structure energy is given by

$$\begin{aligned} U_{band} &= 2\sum_{n=1}^{N_d} \lambda_n g(\lambda_n, \lambda_f) \\ &= 2\sum_{p=1}^{N_d}\sum_{n=1}^{N_d} \lambda_n g(\lambda_n, \lambda_f)|\psi_{n,p}|^2 \\ &= 2\sum_{p=1}^{N_d} \int_a^b \lambda g(\lambda, \lambda_f)\, d\mathcal{E}_{\eta_p,\eta_p}. \end{aligned} \tag{4.45}$$

4.3.1.4 Entropy

The entropy is given by

$$\begin{aligned} S &= -2k_B \sum_{n=1}^{N_d} (g(\lambda_n, \lambda_f)\log g(\lambda_n, \lambda_f) + (1 - g(\lambda_n, \lambda_f))\log(1 - g(\lambda_n, \lambda_f)) \\ &= -2k_B \sum_{p=1}^{N_d}\sum_{n=1}^{N_d} (g(\lambda_n, \lambda_f)\log g(\lambda_n, \lambda_f) + (1 - g(\lambda_n, \lambda_f))\log(1 - g(\lambda_n, \lambda_f))|\psi_{n,p}|^2 \\ &= -2k_B \sum_{p=1}^{N_d} \int_a^b g(\lambda, \lambda_f)\log g(\lambda, \lambda_f) + (1 - g(\lambda, \lambda_f)\log(1 - g(\lambda, \lambda_f)))\, d\mathcal{E}_{\eta_p,\eta_p}. \end{aligned} \tag{4.46}$$

4.3.2 Algorithm

In this section, we describe the proposed LSSGQ method. We solve the nonlinear DFT eigenvalue problem using the SCF method (cf. e.g. Martin (2004)). In each iteration of the SCF method, the electron density is evaluated by solving the linear eigenvalue problem

$$\mathcal{H}\psi_n = \lambda_n \psi_n\,, \quad n = 1, 2, \ldots \tag{4.47}$$

for the eigenvalues and eigenfunctions. The electron density is subsequently evaluated using Eqn. (4.2). However, we circumvent the evaluation of the wavefunctions and directly evaluate the electron density by employing the procedure which we now describe.

We start with the integral representations of the Fermi energy and electron density given in Sections 4.3.1.1 and 4.3.1.2, respectively. We evaluate these integrals using Gauss quadrature described in Section 4.2.2. However, only the operator $\mathcal{H}$ is known and its resolution of the identity $\mathcal{E}(\lambda)$ is unknown a priori. To overcome this, we use the spectral theorem to rewrite the recurrence relation given by Eqn. (4.37) as a Lanczos type iteration

$$\begin{aligned} b_{k+1}v_{k+1} = (\mathcal{H} - a_{k+1})v_k - b_k v_{k-1}\,, \quad k = 0, 1, \ldots, K-1 \\ v_{-1} = 0\,, \quad v_0 = \eta\,, \quad b_0 = 1 \end{aligned} \tag{4.48}$$

where

$$a_{k+1} = (\mathcal{H}v_k, v_k)\,, \quad k = 0, 1, \ldots, K-1 \tag{4.49}$$

and b_k is computed such that $\|v_k\| = 1\,,\ k = 0, 1, \ldots, K-1$. Note that the initial condition $v_0 = \eta$ is chosen depending on the measure $\mathcal{E}_{\eta,\eta}(\lambda)$ with respect to which integration needs to be performed. Thereafter, the nodes and corresponding weights of the quadrature rule can be ascertained following the procedure described in Section 4.2.2. Therefore

$$N_e = 2\sum_{p=1}^{N_d} \int_a^b g(\lambda, \lambda_f)\, d\mathcal{E}_{\eta_p,\eta_p}(\lambda) \approx 2\sum_{p=1}^{N_d}\sum_{k=1}^{K} w_k^{\eta_p} g(\lambda_k^{\eta_p}, \lambda_f) \tag{4.50}$$

$$\rho(\mathbf{x}_0) = 2\int_a^b g(\lambda, \lambda_f)\, d\mathcal{E}_{\zeta,\zeta}(\lambda) \approx 2\sum_{k=1}^{K} w_k^{\zeta} g(\lambda_k^{\zeta}, \lambda_f) \tag{4.51}$$

where $\zeta = \sum_p \eta_p(\mathbf{x}_0)\eta_p$.

Once the self consistent solution of the Kohn-Sham problem for a given position of the nuclei is calculated, the free energy (Eqn. (4.19)) can be calculated by using Gauss quadrature for the evaluation of the band structure energy and entropy

$$\begin{aligned} U_{band} &= 2\sum_{p=1}^{N_d}\int_a^b \lambda g(\lambda,\lambda_f)\, d\mathcal{E}_{\eta_p,\eta_p}(\lambda) \approx 2\sum_{p=1}^{N_d}\sum_{k=1}^{K} w_k^{\eta_p}\lambda_k^{\eta_p} g(\lambda_k^{\eta_p},\lambda_f)\,, \\ S &= -2k_B\sum_{p=1}^{N_d}\int_a^b g(\lambda,\lambda_f)\log g(\lambda,\lambda_f) + (1-g(\lambda,\lambda_f)\log(1-g(\lambda,\lambda_f)))\, d\mathcal{E}_{\eta_p,\eta_p} \\ &\approx -2k_B\sum_{p=1}^{N_d}\sum_{k=1}^{K} w_k^{\eta_p} g(\lambda_k^{\eta_p},\lambda_f)\log g(\lambda_k^{\eta_p},\lambda_f) + (1-g(\lambda_k^{\eta_p},\lambda_f)\log(1-g(\lambda_k^{\eta_p},\lambda_f))). \end{aligned} \tag{4.52}$$

To calculate the ground state free energy, we need to further minimize the free energy with respect to the positions of the nuclei. For this, we equilibriate the forces on the nuclei given by Eqn. (4.17), which is consistent with the energy if the solution is stationary with respect to the electron density and the electrostatic potential. For clarity, we summarize the LSSGQ method in Algorithm 1.

4.3.3 Scaling and performance

In the proposed method, the quantities of interest, namely the Fermi energy, electron density, band structure energy and the entropy are evaluated by performing Gaussian quadrature on their respective integral representations. The cost of evaluating each integral is determined by the calculation of the matrix $\hat{J}_K$ (Eqn. (4.39)) via the recurrence relation (Eqn. (4.48)) and the evaluation of its

Algorithm 1: LSSGQ method

```
Generate guess for electron density (ρ)
repeat Relaxation of atoms
    Calculate charge density of the nuclei (b)
    repeat Self-Consistent loop: SCF
        Calculate electrostatic potential (φ) by solving Eqn. (4.9)
        Calculate exchange correlation potential (V_xc)
        Calculate Fermi-level (λ_f) using Eqn. (4.50)
        Calculate electron density (ρ) using Eqn. (4.51)
        Update the electron density (mixing)
    until Convergence of self-consistent iteration;
    Calculate the forces on the nuclei using Eqn. (4.17)
until Energy minimized with respect to positions of atoms;
Evaluate free energy F using Eqns. (4.52), (4.53) for the band structure energy, entropy
respectively.
```

eigenvalues and eigenvectors. If a localized basis of H is used, such that the matrix representation of $\mathcal{H}$ is sparse, it is clear that the evaluation of the nodes and weights for each integral is independent of the size of the system. Since the number of integrals which need to be evaluated scales linearly with the number of atoms, the entire method has an $\mathcal{O}(M)$ scaling. This is indeed verified by the results presented in Section 4.4.

Now, we discuss the performance of the LSSGQ method. Consider any integral of the form given by Eqn. (4.32). The error incurred by using Gauss quadrature is given by

$$R[f] = \|p_K\|_\zeta^2 \frac{f^{2K}(c)}{(2K)!} \tag{4.53}$$

for some $c \in [a, b]$. Here, K denotes the number of nodes used for the quadrature. It is clear that the accuracy of the spectral quadrature improves with increasing the temperature θ. Again, this is verified by the results presented in Section 4.4.

4.4 Numerical Examples and Validation

In this section, we validate the LSSGQ method through a number of examples. First, we start with a one-dimensional model proposed by García-Cervera et al. (2009) which shares the most important features of a linearized Kohn-Sham problem and therefore can be used to systematically test the accuracy and performance of the LSSGQ method. Subsequently, we utilize the LSSGQ method to solve the nonlinear three-dimensional Kohn-Sham problem in a periodic setting to evaluate the crystal properties. Finally, we show the versatility of the proposed method by studying the phenomenon of surface relaxation.

4.4.1 One-dimensional model

Consider a one-dimensional chain of atoms positioned with unit spacing at $\{R_J\}_{J=1}^M$. Let the effective potential due to an atom at R_J be given by (García-Cervera et al. (2009))

$$V_J(x) = -\frac{\alpha}{\sqrt{2\pi\beta^2}} \exp - \left(\frac{(x - R_J)^2}{2\beta^2} \right) \tag{4.54}$$

and therefore the total potential at any point is given by $V(x) = \sum_{J=1}^M V_J(x)$. Here, α and β represent the depth and width of the wells respectively. By appropriately tuning these parameters, the size of the band gap can be varied, resulting in either metallic, semiconducting or insulating behavior (García-Cervera et al. (2009)). The ground state properties are obtained by solving

$$\mathcal{H}\psi_n = \lambda_n \psi_n \,, \quad n = 1, 2 \ldots \tag{4.55}$$

where the Hamiltonian is given by

$$\mathcal{H} = -\frac{1}{2}\frac{d^2}{dx^2} + V(x). \tag{4.56}$$

The Fermi energy, electron density and energy are evaluated using

$$N_e = 2\sum_{n=1}^{N_d} g(\lambda_n, \lambda_f)\,, \tag{4.57}$$

$$\rho(x) = 2\sum_{n=1}^{N_d} g(\lambda_n, \lambda_f)|\psi_n(x)|^2\,, \tag{4.58}$$

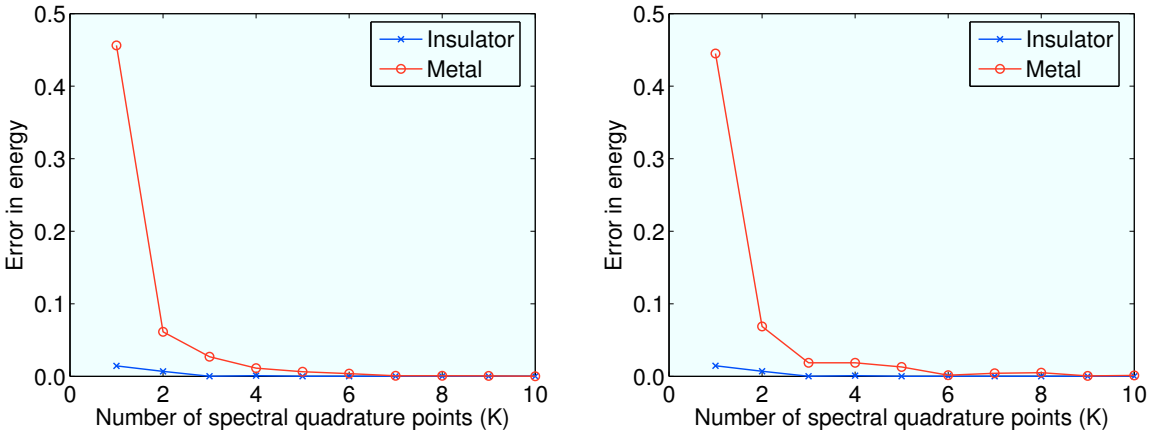
$$\mathcal{E}_\sigma = 2\sum_{n=1}^{N_d} g(\lambda_n, \lambda_f)\lambda_n. \tag{4.59}$$

We solve the problem using the finite-difference approximation with a 12^{th} order accurate discretization for the Laplacian. Details of the LSSGQ method using the finite-difference approximation scheme can be found in Appendix D. By appropriately chosing the parameters α and β, we can vary the band gap of the system, thereby choosing between insulating, semiconducting and metallic systems. Further details can be found in García-Cervera et al. (2009). Here, we consider two sets of parameters $(\alpha, \beta) = (100, 0.3)$ and $(\alpha, \beta) = (10, 0.45)$, which we designate as insulator and metal, respectively. We also study the effect of temperature on the efficacy of the method, by considering two extreme cases corresponding to $\sigma = 0.0001$ and $\sigma = 1.0$. In the results presented below, the error in energy is defined as the difference between the energies obtained by the LSSGQ method and diagonalization. Similarly, error in electron density is the L^2 norm of the difference in electron densities obtained on using the LSSGQ method and diagonalization. Note that all errors

have been normalized with the quantities obtained by diagonalization.

4.4.1.1 Non-periodic calculations

The convergence of the LSSGQ method with respect to the number of spectral quadrature points for both metals and insulators is presented in Fig. 4.1. It is clear that there is rapid convergence of the energy with respect to the number of spectral quadrature points for both metals and insulators. For insulators, the convergence is independent of temperature, a consequence of the presence of a band gap. On the contrary, convergence for metals is significantly accelerated on increasing the temperature.

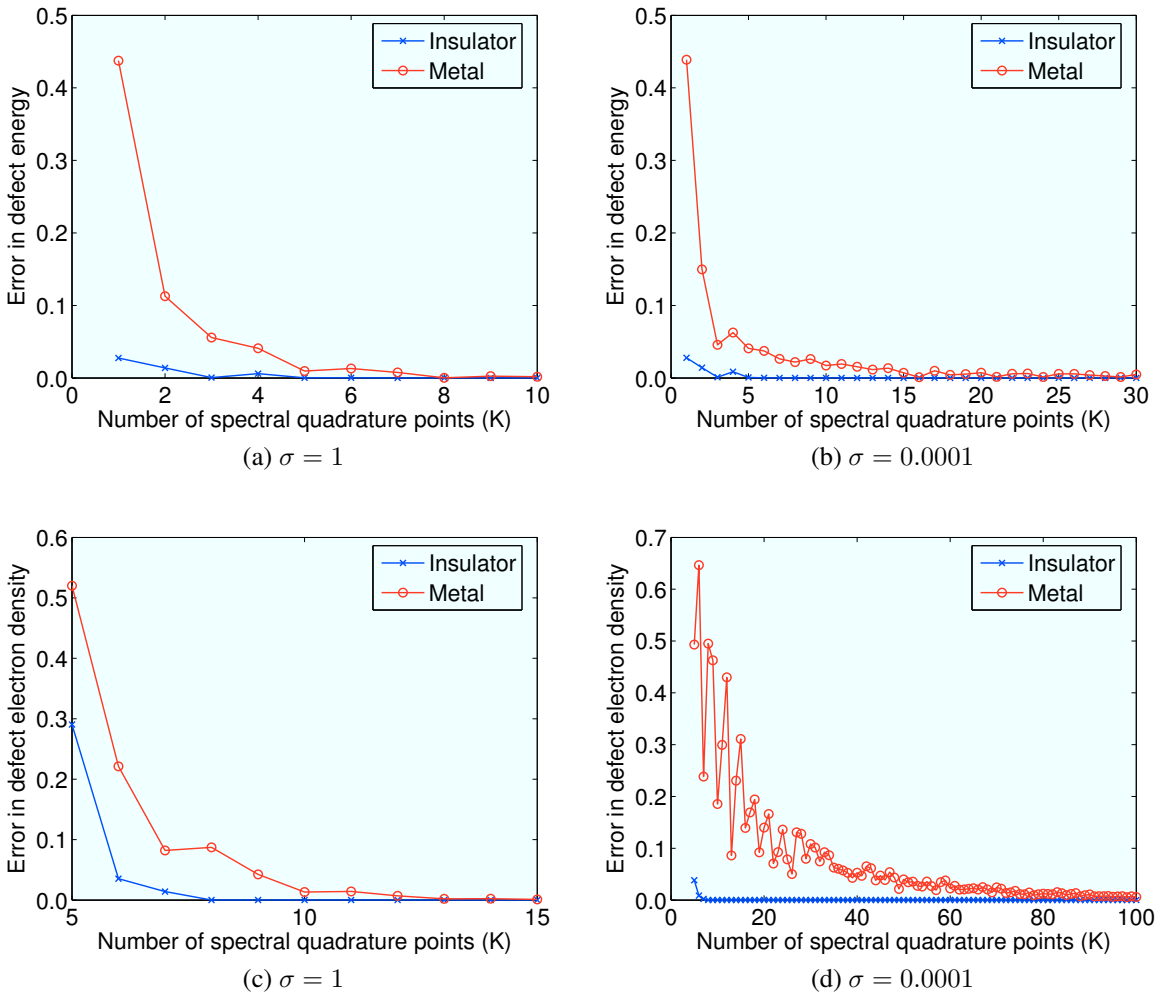


(a) $\sigma = 1$ (b) $\sigma = 0.0001$

Figure 4.1: Convergence of the LSSGQ method

Next, we introduce a defect by removing the center atom and evaluate the defect energy as well as the defect electron density. The defect energy and defect electron density are defined as the difference in energies and electron densities between the systems with and without the defect. Accurately predicting these quantities is challenging, as it requires the calculation of relatively

small differences. The results so obtained have been presented in Fig. 4.2. It is clear that for an insulator, there is rapid convergence with number of spectral quadrature points and independence from temperature. However, for the metal we notice that the number of spectral quadrature points required is dependent on the temperature, with fewer points required on increasing the temperature.



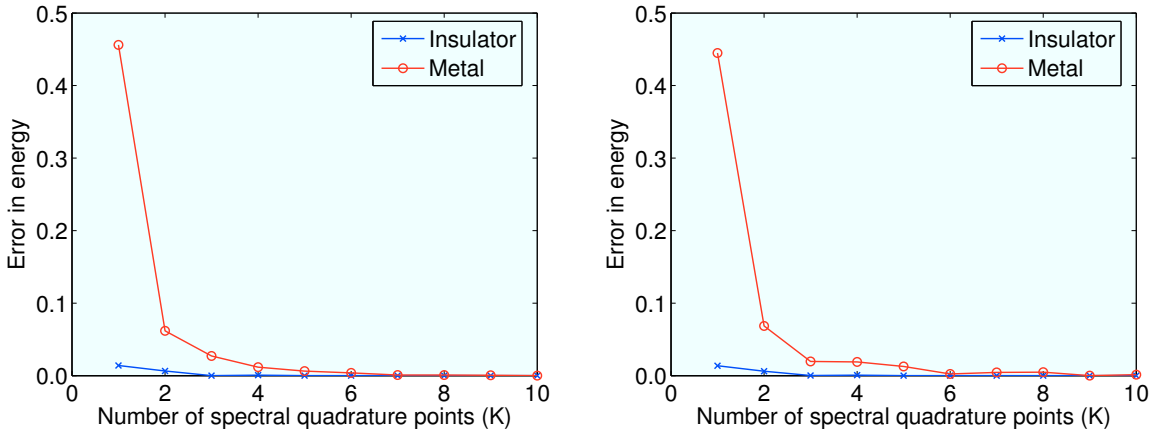
(a) $\sigma = 1$ (b) $\sigma = 0.0001$

(c) $\sigma = 1$ (d) $\sigma = 0.0001$

Figure 4.2: Convergence of the defect energy and defect electron density

4.4.1.2 Periodic calculations

Consider an infinite chain of atoms with unit spacing. We define a two atom unit cell as the representative volume Ω_{RV}. Since the potential decays exponentially, we evaluate the potential inside Ω_{RV} by considering the contribution of atoms which are within a cutoff radius. We then follow the procedure outlined in Appendix E to evaluate the Fermi energy, electron density and energy per atom. The results so obtained are presented in Fig. 4.3. As observed in the non-periodic calculations, convergence in insulators is independent of temperature, whereas for metals convergence is significantly accelerated with increasing temperature.



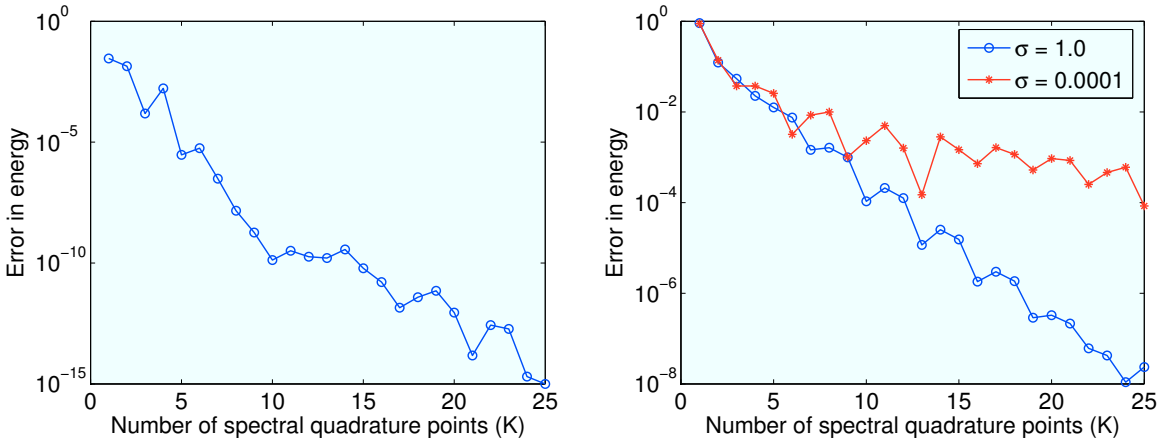
(a) $\sigma = 1$ (b) $\sigma = 0.0001$

Figure 4.3: Convergence of LSSGQ method for the periodic problem

4.4.1.3 Performance and Scaling

Now, we look at the rate of convergence as well as scaling of the proposed LSSGQ method. From Fig. 4.4, it is clear that we obtain spectral convergence with respect to the number of spectral quadrature points. This is to be expected as Gauss quadrature is a spectrally convergent method.

Finally, we present the computational time as a function of the number of atoms in Fig. 4.5. It validates the linear scaling nature of the proposed method.



(a) Insulator (b) Metal

Figure 4.4: Spectral convergence of LSSGQ method

4.4.2 Kohn-Sham problem

In the previous section, we validated the proposed LSSGQ method by applying it to a linear eigenvalue problem representing a one-dimensional chain of atoms. Here, we use the LSSGQ method to solve the three-dimensional nonlinear Kohn-Sham problem using the local density approximation (LDA — Kohn and Sham (1965)) for the exchange correlation energy and the 'Evanescent Core' pseudopotential approximation (Fiolhais et al. (1995)). Specifically, we evaluate the bulk properties of body centered cubic (BCC) crystals of sodium and lithium, and study the phenomenon of surface relaxation in sodium.

We utilize a 6^{th}-order accurate finite-difference discretization of the Laplacian. We designate a representative volume (Ω_{RV}) and utilize the procedure outlined in Appendix E to evaluate the

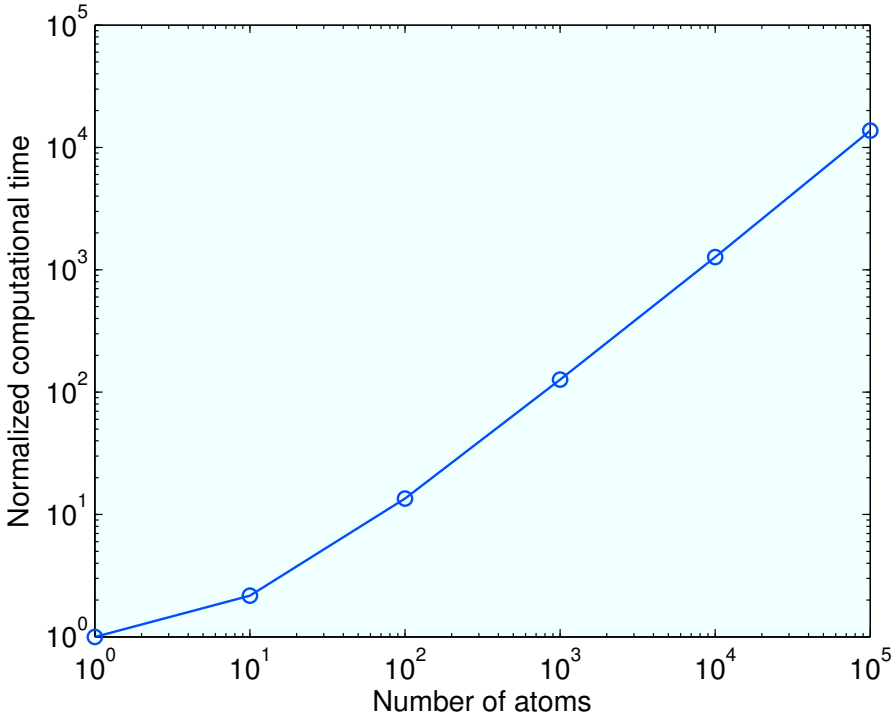


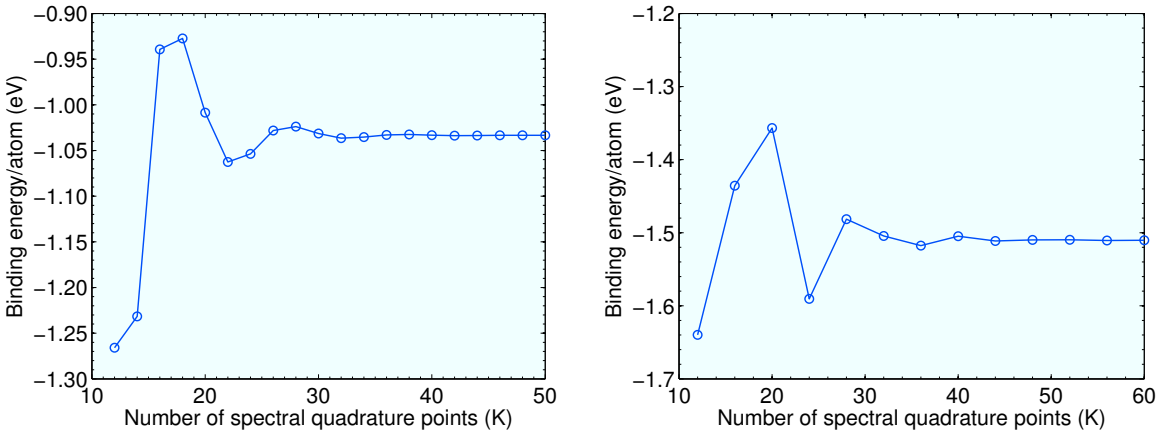
Figure 4.5: Linear scaling of the LSSGQ method

Fermi energy, electron density, band structure energy and entropy. For calculating the Fermi energy λ_f, we solve Eqn. (4.50) using a combination of bisection, secant and inverse quadratic interpolation methods (Forsythe et al. (1973)). We solve the Poisson equation (Eqn. (4.9)) using the generalized minimal residual method (GMRES — Saad and Schultz (1986)). We evaluate the charge density of the nuclei ($b(\mathbf{x}, \mathbf{R})$) by considering the contribution of all nuclei within a prespecified cutoff radius. Such an approximation is extremely accurate owing to the exponential decay of the charge density of the nucleus. For the SCF method, we use mixing with the generalized Broyden method (Fang and Saad (2009)) for acceleration. In order to evaluate the free energy of the system, spatial integrations are performed by assuming that the value of each quantity is constant in a cube of side h around each finite-difference point, where h is the spacing of the finite-difference nodes. We choose $\sigma = 0.8$ eV and extrapolate the finite-temperature calculations to absolute zero using Eqn. (4.23).

4.4.2.1 Crystal properties

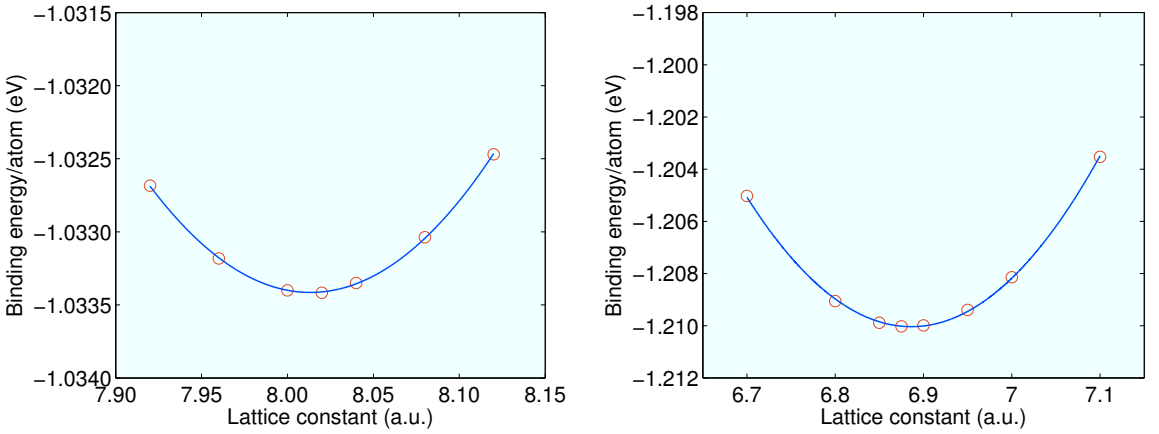
We use the procedure described above to solve for the bulk properties of BCC sodium and lithium. To do so, we designate the BCC unit cell as the representative volume (Ω_{RV}) and utilize periodic boundary conditions. First, we present the convergence in the binding energy per atom with respect to the number of spectral quadrature points (K) in Fig. 4.6. It is clear that we obtain rapid convergence, highlighting the efficacy of the LSSGQ method. Next, we plot the binding energy per atom as a function of lattice constant in Fig. 4.7. Using a cubic fit to this data, we calculate the cohesive energy, equilibrium lattice constant and bulk modulus. The results so obtained are presented in Tables 4.1 and 4.2 with a comparison to previous results obtained by Fiolhais et al.

(1995). The agreement is good, thereby validating the formulation and implementation.



(a) Sodium (b) Lithium

Figure 4.6: Convergence of LSSGQ method



(a) Sodium (b) Lithium

Figure 4.7: Binding energy as a function of lattice constant

4.4.2.2 Surface relaxation

We now study the (001) surface relaxation of BCC sodium. We choose the representative volume (Ω_{RV}) to be a rectangular cuboid of dimensions $a \times a \times h$, where 'a' is the lattice constant. Further,

Table 4.1: Crystalline properties of BCC sodium

Property	LSSGQ	Fiolhais et al. (1995)	Experiment (Fiolhais et al. (1995))
Cohesive energy (eV/atom)	-1.0334	-1.02	-1.04
Lattice constant (a.u.)	8.01	8.21	8.21
Bulk modulus (GPa)	5.0	7.1	7.3

Table 4.2: Crystalline properties of BCC lithium

Property	LSSGQ	Fiolhais et al. (1995)	Experiment (Fiolhais et al. (1995))
Cohesive energy (eV/atom)	-1.21	-1.31	-1.0
Lattice constant (a.u.)	6.87	6.77	6.77
Bulk modulus (GPa)	10.0	14.0	13.0

$h = h_{vac} + h_{cell}$, where $h_{vac} = n_1 a$ and $h_{cell} = n_2 a$ ($n_1, n_2 \in \mathbb{Z}$) are the heights of vaccuum and material included in Ω_{RV}. We use periodic boundary conditions in the x_1, x_2 directions. Further, we use zero Dirichlet boundary conditions on the top face of Ω_{RV} and assume that the perfectly periodic solution is attained in the bottom most unit cell of Ω_{RV}. We relax the atoms only along the x_3 direction and their minimum configuration is obtained using the BFGS quasi-Newton method with a cubic line search procedure (cf. e.g. Vogel (2002)). We evaluate the surface energy using the relation

$$\mathcal{E}_{surf} = \frac{\mathcal{E}_{\Omega_{RV}} - N_e \mathcal{E}_{coh}}{a^2} \tag{4.60}$$

where $\mathcal{E}_{\Omega_{RV}}$ is the total energy in Ω_{RV}, $\mathcal{E}_{coh}$ is the cohesive energy for a perfect crystal and N_e is the number of electrons in Ω_{RV}.

First, we verify the convergence of the calculations with respect to the number of spectral quadrature points (K) in Fig. 4.8. It is clear that we obtain rapid convergence, highlighting the efficacy of the method. Next, we present the calculated surface energy and the displacement of the

atoms in Table 4.3. Note that the forces on the atoms in the third layer and beyond are below the threshold value used for the calculations. Finally, we plot the electron density contours on the mid plane and edge plane in Figs. 4.9 and 4.10, respectively.

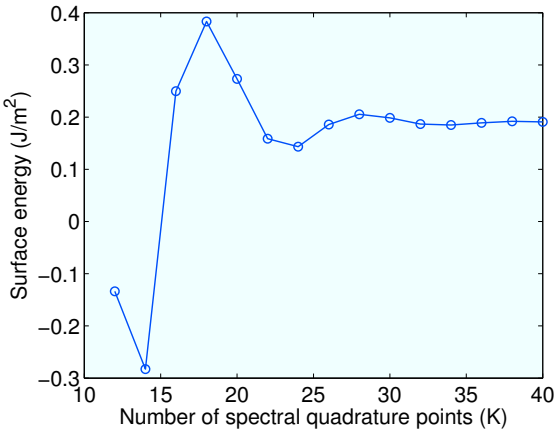


Figure 4.8: Convergence of LSSGQ method for the surface problem

Table 4.3: Properties of the (001) surface of BCC sodium

Surface energy	0.2 J/m^2
Displacement of atoms	
First layer	0.75 a.u.
Second layer	-0.19 a.u.

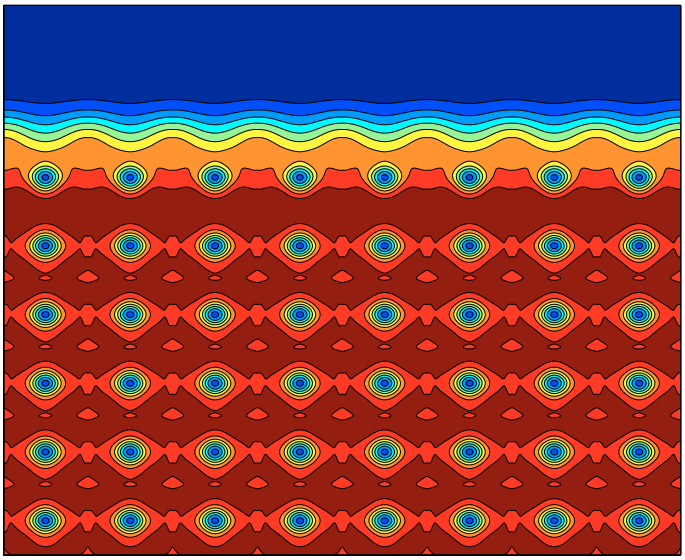


Figure 4.9: Electron density contours on the mid plane of a surface of (001) BCC sodium

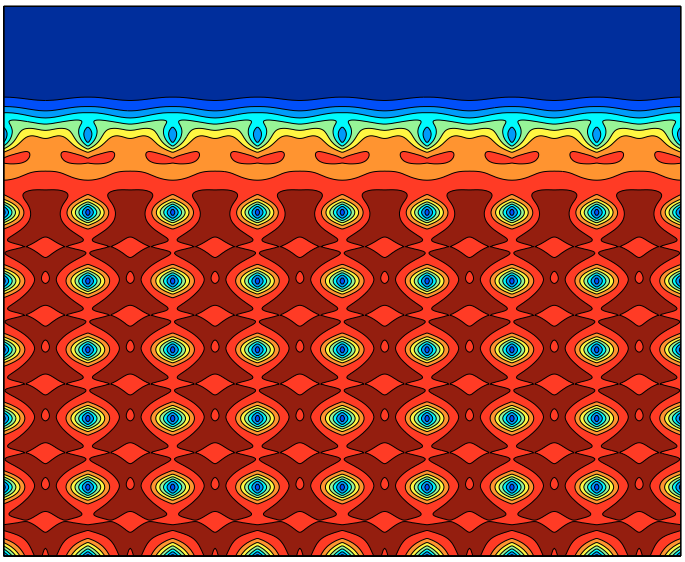


Figure 4.10: Electron density contours on the face plane of a surface of (001) BCC sodium

Chapter 5

Coarse-grained Kohn-Sham density functional theory

In spite of the recent advances in DFT, the study of crystal defects at realistic concentrations remains intractable. Defects present a unique challenge since both the electronic structure of the core as well as the long range elastic field need to be simultaneously resolved. Multiscale approaches which coarse-grain DFT or embed it into simpler, less accurate models like tight-binding (TB) or empirical potentials have also been developed recently (Choly et al. (2005); Lu et al. (2006); Govind et al. (1999); Bernstein et al. (2009)). Though these multiscale methods provide valuable insight, they suffer from a few notable drawbacks. In some cases, there is no seamless transition from DFT to TB or empirical potentials, while in others, uncontrolled approximations made by the use of linear response theory or Cauchy Born hypothesis render them unreliable. Also, there is no systematic convergence of the solution of these models to the full DFT solution.

In this chapter, we present a formulation to seamlessly coarse-grain DFT (CG-DFT) solely based on approximation theory, without the introduction of any new equations or physics. This allows us to study defects at a fraction of the original computational cost, without any appreciable loss of accuracy. Additionally, we can obtain systematic convergence to the fully resolved solution.

Below, we discuss the formulation of CG-DFT in Section 5.1 and validate it throught examples in Section 5.2. Note that CG-DFT builds on the linear scaling method (LSSGQ) presented in Chapter 4, which should be read first to understand the description provided here.

5.1 Formulation

The central idea of CG-DFT is as follows. When a defect is introduced in an otherwise perfect lattice, the resultant perturbation is expected to be localized in the vicinity of the defect and decays to zero as one moves away from the defect. We incorporate this intuition into our numerical methodology as described below. For simplicity and clarity of the key ideas involved, we discuss the method in terms of the finite-difference approximation scheme, which is used to validate the formulation in Section 5.2. It is relatively simple to extend these ideas to other choice of bases.

In electronic structure calculations, the main quantities of interest are electron density, electrostatic potential, the positions of the nuclei and the energy of the system. We introduce coarse-graining type approximations, which allows the evaluation of these quantities at reduced computational cost. First, we discuss the evaluation of the electron density, which constitutes the majority of the computational time. Let us denote the domain of interest by Ω, initially spanned by a perfect lattice and discretized using finite-difference nodes, the collection of whom we denote by $\mathcal{N}$. First, we perform a periodic calculation for the perfect lattice using the LSSGQ method and map this solution onto $\mathcal{N}$. Next, we introduce the defect and choose a collection of representative finite-difference nodes, which we denote by $\mathcal{R}$. The nodes in $\mathcal{R}$ are chosen such that there is a high concentration of nodes in the vicinity of the defect and progressively dilutes with increasing

distance from the defect. For the LSSGQ method with the finite-difference approximation, we need to evaluate the spectral quadrature nodes and weights for all finite-difference nodes $p \in \mathcal{N}$ (See Appendix D). However, in CG-DFT we only calculate the spectral nodes and weights for the finite-difference nodes $p \in \mathcal{R}$. Then, we introduce the following approximation for the electron density

$$\rho^p = \rho^p_{per} + \sum_{q \in \mathcal{R}} \gamma^p_q (\rho^q - \rho^q_{per}) \tag{5.1}$$

where ρ_{per} is the corresponding periodic solution and γ^q_p are the weights such that $\gamma^q_p = 1$ if $p = q$, else it is decided by the degree of interpolation/approximation used. The calculation of the Fermi energy proceeds as before by solving for the constraint

$$N_e = h^D \sum_{p=1}^{N_d} \rho^p. \tag{5.2}$$

Additionally, we introduce the following approximations

$$u^p = u^p_{per} + \sum_{q \in \mathcal{R}} \gamma^p_q (u^q - u^q_{per}) \tag{5.3}$$

$$s^p = s^p_{per} + \sum_{q \in \mathcal{R}} \gamma^p_q (s^q - s^q_{per}) \tag{5.4}$$

which allow the calculation of the band structure energy and entropy as before

$$U_{band} = \sum_{p=1}^{N_d} u^p \tag{5.5}$$

$$S = \sum_{p=1}^{N_d} s^p. \tag{5.6}$$

The electrostatic potential is evaluated as follows

$$\phi = \phi_{per} + \phi_d \tag{5.7}$$

where ϕ_{per} is the periodic solution and ϕ_d is calculated by solving the equation

$$-\frac{1}{4\pi}\nabla^2\phi_d = \rho - \rho_{per}. \tag{5.8}$$

By its very nature, ϕ_d is localized near the defect and decays to zero away from it. Therefore, it can be easily calculated by a discretization adapted to this feature. Finally, for finding the equilibrium position of the atoms, we select representative atoms and interpolate the displacement for the remaining atoms as done traditionally in the quasi-continuum method (cf. e.g. Tadmor et al. (1996); Knap and Ortiz (2001)). Again, the representative atoms are chosen such that there is a high concentration in the vicinity of the defect which reduces with increasing distance from the defect.

5.2 Validation

In this section, we validate the proposed CG-DFT method. We study a one-dimensional model problem in Section 5.2.1 and a Kohn-Sham surface relaxation problem in Section 5.2.2.

5.2.1 One-dimensional model

We study the efficacy of CG-DFT in the framework of the one-dimensional model presented in Section 4.4.1. We use a 12^{th} order accuracte finite-difference appeoximation of the Laplacian. Consider a chain constituting of 101 atoms with unit spacing. First, we introduce a defect by removing the center atom. Next, we coarse-grain in both directions from the position of the defect and adopt the procedure described in Section 5.1 to solve the problem. We plot the convergence in the defect energy and defect electron density as a function of the number of representative nodes in Fig. 5.1 respectively. The error is measured and normalized with respect to the fully resolved LSSGQ solution. It is clear that we get rapid convergence in the solution for both metals and insulators. Convergence in insulators is much more rapid and this is to be expected because of the localized nature of defects in insulators. Therefore, we conclude that using CG-DFT, we can solve the defect problem with a small fraction of the original computational cost.

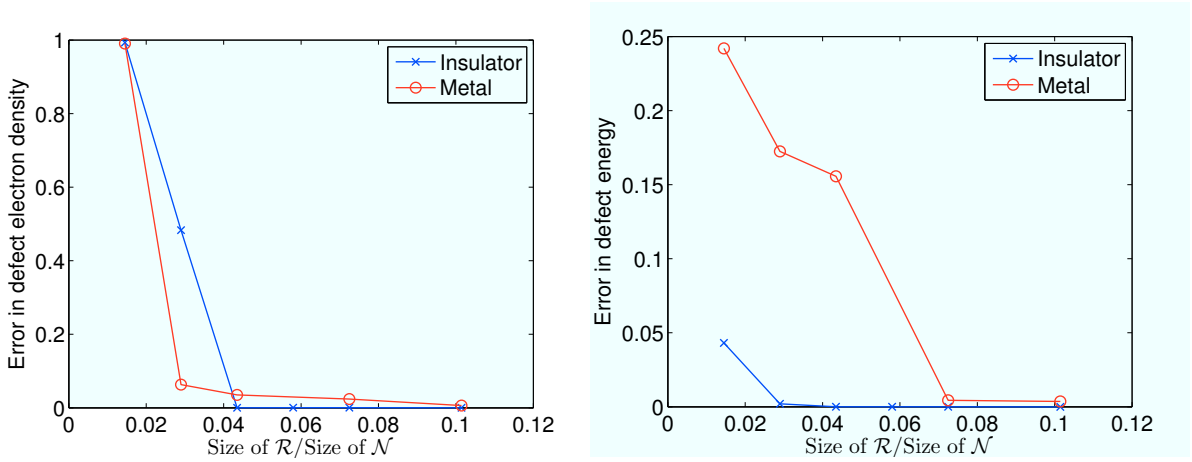


Figure 5.1: Convergence of the CG-DFT method

5.2.2 Kohn-Sham problem—surface relaxation

We now study the (001) surface relaxation of BCC sodium using CG-DFT. We employ the procedure described in Section 4.4.2 with the introduction of coarse-graining described in Section 5.1. We introduce coarse-graining for the atomic positions, electron density and local band structure energy. Note that we do not introduce coarse-graining for the electrostatic potential since the computational time involved for its calculation is negligible. We compare the results obtained using CG-DFT with the fully resolved solution obtained using the LSSGQ method. A summary of the results is presented in Table 5.1. We can see that there are significant computational savings when using CG-DFT, with no appreciable loss of accuracy. It is worth noting that the surface relaxation problem is effectively one-dimensional. Therefore, coarse-graining can be performed only in one direction. As a consequence, the computational savings achieved are not as significant as is possible in a truly two/three dimensional problem. We expect that when CG-DFT is used to study other defects like vacancies, dislocations and cracks, we would achieve far more substantial savings.

Table 5.1: Summary of the results for CG-DFT

Number of representative nodes/Total number of nodes	0.25
Number of representative atoms/Total number of atoms	0.3
Normalized error in surface energy	0.12 %
Normalized error in displacement of atoms	
First layer	0.11 %
Second layer	0.17 %

Chapter 6

Conclusions

We have developed a real-space, non-periodic finite-element formulation of DFT. By including the electrostatic potential among the unknown fields we have reformulated the problem as a local saddle-point problem. We have shown the well-posedness of this formulation for both the all-electron problem as well as the pseudopotential approximation. In particular, we have proved the existence of solutions, and, in addition, the convergence of finite-element approximations, including numerical quadratures, using Γ-convergence methods. We have also developed a parallel implementation of this formulation capable of performing both all-electron and pseudopotential calculations. In this implementation, the advantageous features of both the SCF and Newton's methods have been combined to ensure rapid convergence to the solution. The flexibility provided by the unstructured nature of the finite-element method is exploited by optimally coarsening the triangulation away from the nuclei and by convecting the triangulation with the atomic positions, thereby enhancing the efficiency of the calculations without loss of accuracy. The formulation has been tested through a number of examples and the accuracy of the results is in accord with the literature. Also, various aspects of the numerical performance of the implementation have been investigated, including the convergence rate of the finite-element method and its scaling with

increasing system size and number of processors. We obtain ostensibly ideal convergence rate for the finite-element method, linear or $O(N)$-scaling with problem size and good parallel scalability.

A shortcoming of low-order finite-elements in general, is the large number of basis functions required to obtain convergence compared to other bases such as plane waves. To mitigate this deficiency, we have developed a non-periodic, real-space, mesh-free convex approximation scheme for Kohn-Sham density functional theory. We have developed a parallel implementation of this formulation capable of performing both all-electron and pseudopotential calculations. The formulation has been tested through a number of examples and the accuracy of the results is in accord with the literature. We have shown that the use of *max-ent* basis functions can result in a significant reduction in the required number of basis functions compared to the corresponding simplicial linear finite-elements. However, this comes at the price of increased non-locality of the basis functions as well as the need for significantly higher order quadrature, which in some cases could outweigh the gains from the requirement of fewer basis functions. Also, as in most mesh-free methods, the necessity of a background mesh for performing numerical integration makes the method not completely mesh free. Further work in these aspects is required to make the method competitive with conventional bases used in electronic structure calculations.

We have also developed a real-space, non-periodic formulation for coarse-graining DFT (CG-DFT). We have achieved this in two stages. First, we have developed a linear-scaling method for DFT (LSSGQ). We know that the wavefunctions are global quantities. This is a consequence of their orthonormality constraint. Therefore, they are not amenable to coarse-graining. To overcome this, we have developed the LSSGQ method where we directly evaluate the electron density without evaluating any of the wavefunctions. Second, we introduce coarse-graining approximations in

the LSSGQ method, whereby we are able to study defects at a fraction of the original computational cost. This has been achieved soley via approximation theory, without the introduction of any equations and spurious physics. By doing so, we have obtained systematic convergence to the fully resolved solution. We have verified both the LSSGQ and CG-DFT methods through examples, for which we have developed higher order finite difference implementations of the same.

This work lays the foundation for analyzing samples (using DFT) of sufficient size to yield meaningful size-independent properties of isolated lattice defects, such as vacancies and dislocations, in concentrations representative of actual material systems. This method would also be useful for studying other defects like interfaces, stacking faults, domain walls and cracks. The insight gained from these studies would be particularly useful in understanding the role of defects on material strength, a question of great scientific and technological importance. We believe that this thesis is a significant step in this direction.

Appendix A

Computation of forces in the finite-element formulation

For completeness, we derive the forces that determine the equilibrium nuclear configuration. To do so, we assume that for any configuration of the nuclei, the variational problem defined by

$$\inf_{\Psi\in(H_0^1(\Omega))^N} \sup_{\phi\in H_0^1(\Omega)} \mathcal{E}(\Psi,\phi,\mathbf{R}) \tag{A.1}$$

subject to the constraints given by Eqn. (2.21), has been solved. The derivation below closely follows Thoutireddy (2002) and Gavini et al. (2007a). Using Lagrange multipliers to enforce the constraints, we have

$$\begin{aligned}\mathcal{E}_c(\Psi,\phi,\mathbf{R},\Lambda^\alpha,\Lambda^\beta) = \int_\Omega f(\Psi,\nabla\Psi)\,\mathrm{d}\mathbf{x} + \int_\Omega \left(\rho(\mathbf{x})+b(\mathbf{x},\mathbf{R})\right)\phi(\mathbf{x},\mathbf{R})\,\mathrm{d}\mathbf{x} \\ -\frac{1}{8\pi}\int_\Omega |\nabla\phi(\mathbf{x},\mathbf{R})|^2\,\mathrm{d}\mathbf{x} - \sum_{i=1}^{N_\alpha}\sum_{j=1}^{N_\alpha}\lambda_{ij}^\alpha\left(\int_\Omega \psi_{i\alpha}^*(\mathbf{x})\psi_{j\alpha}(\mathbf{x})\,\mathrm{d}\mathbf{x}-\delta_{ij}\right) \\ -\sum_{i=1}^{N_\beta}\sum_{j=1}^{N_\beta}\lambda_{ij}^\beta\left(\int_\Omega \psi_{i\beta}^*(\mathbf{x})\psi_{j\beta}(\mathbf{x})\,\mathrm{d}\mathbf{x}-\delta_{ij}\right).\end{aligned} \tag{A.2}$$

In this situation, consider $I_1^h + I_2^h + I_3^h + I_4^h$, where

$$
\begin{aligned}
I_1^h &= \int_\Omega f(\Psi, \nabla\Psi)\,\mathrm{d}\mathbf{x}\,, \\
I_2^h &= \int_\Omega \left(\rho(\mathbf{x}) + b(\mathbf{x},\mathbf{R})\right)\phi(\mathbf{x},\mathbf{R})\,\mathrm{d}\mathbf{x}\,, \\
I_3^h &= -\frac{1}{8\pi}\int_\Omega |\nabla\phi(\mathbf{x},\mathbf{R})|^2\,\mathrm{d}\mathbf{x}\,, \\
I_4^h &= -\sum_{i=1}^{N_\alpha}\sum_{j=1}^{N_\alpha}\lambda_{ij}^{\alpha}\left(\int_\Omega \psi_{i\alpha}^*(\mathbf{x})\psi_{j\alpha}(\mathbf{x})\,\mathrm{d}\mathbf{x} - \delta_{ij}\right) - \sum_{i=1}^{N_\beta}\sum_{j=1}^{N_\beta}\lambda_{ij}^{\beta}\left(\int_\Omega \psi_{i\beta}^*(\mathbf{x})\psi_{j\beta}(\mathbf{x})\,\mathrm{d}\mathbf{x} - \delta_{ij}\right).
\end{aligned}
$$

Note that

$$
I_1^h = \int_\Omega f(\Psi,\nabla\Psi)\,\mathrm{d}\mathbf{x} = \sum_{e\in T_h}\int_{\hat\Omega} f(\Psi,\nabla\Psi)\det\left(\frac{\partial \mathbf{x}_M}{\partial \hat{\mathbf{x}}_N}\right)\mathrm{d}\hat{\mathbf{x}}\,,
$$

where $\hat\Omega$ is the reference volume in isoparametric formulation and $\frac{\partial \mathbf{x}_M}{\partial \hat{\mathbf{x}}_N}$ is the Jacobian of the transformation. Taking first variations of I_1, we obtain

$$
\begin{aligned}
\delta I_1^h &= \sum_{e\in\mathcal{T}_0}\int_{\hat\Omega}\left\{-\sum_\sigma\sum_{i=1}^{N_\sigma}\frac{\delta f(\Psi,\nabla\Psi)}{\delta\psi_{i\sigma,J}}\left[\sum_{a=1}^{N_h}\psi_{i\sigma}^a\hat{N}_{a,A}\frac{\partial\hat{\mathbf{x}}_A}{\partial\mathbf{x}_K}(\sum_{b=1}^{N_h}\delta\mathbf{x}_{bK}^e\hat{N}_{b,B})\frac{\partial\hat{\mathbf{x}}_B}{\partial\mathbf{x}_J}\right]\right. \\
&\qquad \left. + f(\Psi,\nabla\Psi)(\sum_{b=1}^{N_h}\delta\mathbf{x}_{bK}^e\hat{N}_{b,B})\frac{\partial\hat{\mathbf{x}}_B}{\partial\mathbf{x}_K}\right\}\det\left(\frac{\partial\mathbf{x}_M}{\partial\hat{\mathbf{x}}_N}\right)\,\mathrm{d}\hat{\mathbf{x}} \\
&= \sum_{e\in\mathcal{T}_0}\int_{\Omega^e}\left\{-\sum_\sigma\sum_{i=1}^{N_\sigma}\frac{\delta f(\Psi,\nabla\Psi)}{\delta\psi_{i\sigma,J}}\left[\sum_{a=1}^{N_h}\psi_{i\sigma}^a N_{a,K}\right] + f(\Psi,\nabla\Psi)\delta_{KJ}\right\}\left(\sum_{b=1}^{N_h}\delta\mathbf{x}_{bK}^e N_{b,J}\right)\,\mathrm{d}\mathbf{x} \\
&= \sum_{e\in\mathcal{T}_0}\int_{\Omega^e}\left\{-\sum_\sigma\sum_{i=1}^{N_\sigma}\frac{\delta f(\Psi,\nabla\Psi)}{\delta\psi_{i\sigma,J}}\psi_{i\sigma,K}(\mathbf{x}) + f(\Psi,\nabla\Psi)\delta_{KJ}\right\}\left(\sum_{b=1}^{N_h}\delta\mathbf{x}_{bK}^e N_{b,J}\right)\,\mathrm{d}\mathbf{x}\,.
\end{aligned}
$$

Similarly, note that

$$I_2^h = \int_\Omega \left(\rho(\mathbf{x}) + b(\mathbf{x},\mathbf{R})\right)\phi(\mathbf{x},\mathbf{R})\,\mathrm{d}\mathbf{x} = \sum_{e\in\mathcal{T}_0}\int_{\hat{\Omega}} \left(\rho(\mathbf{x}) + b(\mathbf{x},\mathbf{R})\right)\phi(\mathbf{x},\mathbf{R})\det\left(\frac{\partial \mathbf{x}_M}{\partial \hat{\mathbf{x}}_N}\right)\mathrm{d}\hat{\mathbf{x}}\,.$$

Taking variations, we find

$$\begin{aligned}
\delta I_2^h &= \sum_{e\in\mathcal{T}_0}\int_{\hat{\Omega}} \left(\rho(\mathbf{x}) + b(\mathbf{x},\mathbf{R})\right)\phi(\mathbf{x},\mathbf{R})\left(\sum_{b=1}^{N_h}\delta\mathbf{x}_{bK}^e \hat{N}_{b,B}\right)\frac{\partial\hat{\mathbf{x}}_B}{\partial\mathbf{x}_K}\det\left(\frac{\partial \mathbf{x}_M}{\partial \hat{\mathbf{x}}_N}\right)\mathrm{d}\hat{\mathbf{x}} \\
&\quad + \sum_{e\in\mathcal{T}_0}\int_{\hat{\Omega}} (\delta b)\,\phi(\mathbf{x},\mathbf{R})\det\left(\frac{\partial \mathbf{x}_M}{\partial \hat{\mathbf{x}}_N}\right)\mathrm{d}\hat{\mathbf{x}} \\
&= \sum_{e\in\mathcal{T}_0}\int_{\Omega^e} \left(\rho(\mathbf{x}) + b(\mathbf{x},\mathbf{R})\right)\phi(\mathbf{x},\mathbf{R})\delta_{KJ}\left(\sum_{b=1}^{N_h}\delta\mathbf{x}_{bK}^e N_{b,J}\right)\mathrm{d}\mathbf{x} \\
&\quad + \sum_{e\in\mathcal{T}_0}\int_{\Omega^e} b(\mathbf{x},\mathbf{R})\left(\sum_{a=1}^{N_h}\phi_a N_{a,K}\right)\delta\mathbf{x}_{bK}^e\,\mathrm{d}\mathbf{x}\ .
\end{aligned}$$

Similarly,

$$\begin{aligned}
\delta I_3^h &= -\frac{1}{8\pi}\sum_{e\in\mathcal{T}_0}\int_{\Omega^e}\left\{|\nabla\phi(\mathbf{x},\mathbf{R})|^2\delta_{KJ} - 2\phi_{,J}(\mathbf{x},\mathbf{R})\phi_{,K}(\mathbf{x},\mathbf{R})\right\}\left(\sum_{b=1}^{N_h}\delta\mathbf{x}_{bK}^e N_{b,J}\right)\mathrm{d}\mathbf{x}\,, \\
\delta I_4^h &= -\sum_{i=1}^{N_\alpha}\sum_{j=1}^{N_\alpha}\lambda_{ij}^\alpha\sum_{e\in\mathcal{T}_0}\int_{\Omega^e}\psi_{i\alpha}^*(\mathbf{x})\psi_{j\alpha}(\mathbf{x})\delta_{KJ}\left(\sum_{b=1}^{N_h}\delta\mathbf{x}_{bK}^e N_{b,J}\right)\mathrm{d}\mathbf{x} \\
&\quad -\sum_{i=1}^{N_\beta}\sum_{j=1}^{N_\beta}\lambda_{ij}^\beta\sum_{e\in\mathcal{T}_0}\int_{\Omega^e}\psi_{i\beta}^*(\mathbf{x})\psi_{j\beta}(\mathbf{x})\delta_{KJ}\left(\sum_{b=1}^{N_h}\delta\mathbf{x}_{bK}^e N_{b,J}\right)\mathrm{d}\mathbf{x}.
\end{aligned}$$

Collecting all terms, the force on the b^{th} node along the K^{th} direction is given by

$$f^{bK} = \sum_{e\in\mathcal{T}_0}\int_{\Omega^e} Z_b b_b\left(\sum_{a=1}^{N_h}\phi_a N_{a,K}\right)\mathrm{d}\mathbf{x} + \sum_{e\in\mathcal{T}_0}\int_{\Omega^e}\mathrm{E}_{KJ}N_{b,J}\,\mathrm{d}\mathbf{x}\,, \tag{A.3}$$

where

$$
\begin{aligned}
\mathrm{E}_{KJ} \quad &= \left\{ f(\Psi, \nabla\Psi) + \left(\rho(\mathbf{x}) + b(\mathbf{x}, \mathbf{R})\right) \phi(\mathbf{x}, \mathbf{R}) - \frac{1}{8\pi} |\nabla\phi(\mathbf{x}, \mathbf{R})|^2 \right\} \delta_{KJ} \\
- \quad &\left\{ \sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\alpha} \lambda_{ij}^{\alpha} \psi_{i\alpha}^{*}(\mathbf{x}) \psi_{j\alpha}(\mathbf{x}) + \sum_{i=1}^{N_\beta} \sum_{j=1}^{N_\beta} \lambda_{ij}^{\beta} \psi_{i\beta}^{*}(\mathbf{x}) \psi_{j\beta}(\mathbf{x}) \right\} \delta_{KJ} \\
- \quad &\sum_{\sigma} \sum_{i=1}^{N_\sigma} \frac{\delta f(\Psi, \nabla\Psi)}{\delta \psi_{i\sigma,J}} \psi_{i\sigma,K}(\mathbf{x}) + \frac{1}{4\pi} \phi_{,J}(\mathbf{x}, \mathbf{R}) \phi_{,K}(\mathbf{x}, \mathbf{R}) \, . \qquad \text{(A.4)}
\end{aligned}
$$

Note that the first term of Eqn. (A.3) corresponds to the Hellmann-Feynman force on the nuclei, whereas the second term arises because of the finite-element discretization and acts on all the nodes of the triangulation. The latter arises from the need to minimize the energy with respect to the nodal configuration for a fixed number of nodes. Hence it can be used effectively for a posteriori mesh adaption. Note that the discretization force is local.

Appendix B

Spatial derivatives of the Shape Functions

In this section we derive expressions for the spatial derivatives of the basis functions to be used in the mesh-free approximation scheme described in Chapter 3. We define the following quantities from Eqns (3.10) and (3.17):

$$\begin{aligned} f_a(\mathbf{x},\boldsymbol{\lambda},\beta) &= -\beta h_a(|\mathbf{x}-\mathbf{x}_a|)+\boldsymbol{\lambda}.(\mathbf{x}-\mathbf{x}_a)\,, \\ p_a(\mathbf{x},\boldsymbol{\lambda},\beta) &= \frac{\exp(f_a(\mathbf{x},\boldsymbol{\lambda},\beta))}{\sum_{b=1}^{P}\exp(f_b(\mathbf{x},\boldsymbol{\lambda},\beta))}\,, \\ \mathbf{r}(\mathbf{x},\boldsymbol{\lambda},\beta) &= -\frac{\partial\hat{g}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}=\sum_{a=1}^{P}p_a(\mathbf{x},\boldsymbol{\lambda},\beta)(\mathbf{x}-\mathbf{x}_a)\,, \\ J(\mathbf{x},\boldsymbol{\lambda},\beta) &= -\frac{\partial^2\hat{g}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}^2}=\sum_{a=1}^{P}p_a(\mathbf{x},\boldsymbol{\lambda},\beta)(\mathbf{x}-\mathbf{x}_a)\otimes(\mathbf{x}-\mathbf{x}_a)-\mathbf{r}(\mathbf{x},\boldsymbol{\lambda},\beta)\otimes\mathbf{r}(\mathbf{x},\boldsymbol{\lambda},\beta). \end{aligned} \tag{B.1}$$

Given a function $h(\mathbf{x},\boldsymbol{\lambda},\beta)$, we define $h^* = h(\mathbf{x},\boldsymbol{\lambda}^*,\beta)$, where λ^* is given by Eqn. (3.12). It follows that

$$\nabla p_a^* = p_a^*\left(\nabla f_a^* - \sum_{b=1}^{P} p_b^*\nabla f_b^*\right). \tag{B.2}$$

Using the chain rule we have

$$\nabla f_a^* = \left(\frac{\partial f_a}{\partial \mathbf{x}}\right)^* + \left(\frac{\partial f_a}{\partial \boldsymbol{\lambda}}\right)^* D\boldsymbol{\lambda}^* \tag{B.3}$$

where

$$\left(\frac{\partial f_a}{\partial \mathbf{x}}\right)^* = -\beta h_a^{'} \frac{(\mathbf{x}-\mathbf{x}_a)}{|\mathbf{x}-\mathbf{x}_a|} + \boldsymbol{\lambda}^*(\mathbf{x})\,, \tag{B.4}$$

$$\left(\frac{\partial f_a}{\partial \boldsymbol{\lambda}}\right)^* = \mathbf{x}-\mathbf{x}_a. \tag{B.5}$$

To evaluate $D\boldsymbol{\lambda}^*$, we note

$$D\mathbf{r}^* = \left(\frac{\partial \mathbf{r}}{\partial \mathbf{x}}\right)^* + \left(\frac{\partial \mathbf{r}}{\partial \boldsymbol{\lambda}}\right)^* D\boldsymbol{\lambda}^* = 0 \tag{B.6}$$

where

$$\left(\frac{\partial \mathbf{r}}{\partial \boldsymbol{\lambda}}\right)^* = J^*\,, \tag{B.7}$$

$$\left(\frac{\partial \mathbf{r}}{\partial \mathbf{x}}\right)^* = -\beta \sum_{a=1}^{P} p_a^* h_a^{'} \frac{(\mathbf{x}-\mathbf{x}_a)\otimes(\mathbf{x}-\mathbf{x}_a)}{|\mathbf{x}-\mathbf{x}_a|} + I_d \equiv K. \tag{B.8}$$

I_d denotes the identity matrix. It follows

$$\nabla f_a^* = -\beta h_a^{'} \frac{(\mathbf{x}-\mathbf{x}_a)}{|\mathbf{x}-\mathbf{x}_a|} + \boldsymbol{\lambda}^*(\mathbf{x}) - (\mathbf{x}-\mathbf{x}_a) J^{*-1} K \tag{B.9}$$

and

$$\nabla p_a^* = p_a^* \left[\beta \sum_{b=1}^{P} p_b^* h_b^{'} \frac{(\mathbf{x} - \mathbf{x}_b)}{|\mathbf{x} - \mathbf{x}_b|} - \beta h_a^{'} \frac{(\mathbf{x} - \mathbf{x}_a)}{|\mathbf{x} - \mathbf{x}_a|} - (\mathbf{x} - \mathbf{x}_a) J^{*-1} K \right]. \tag{B.10}$$

Appendix C

Padé approximation and recursion method

The recursion method (Haydock (1980)) has found many applications in the physics literature. Here, we discuss the close connection between the LSSGQ and recursion methods. In doing so, we also establish their relationship to the Padé approximation. Details of the notation not described here can be found in Section 4.

Let $\mathcal{R}_{\mathcal{H}}(z) = (z\mathcal{I} - \mathcal{H})^{-1}$ represent the resolvent of the operator $\mathcal{H}$ for $z \in \rho(\mathcal{H}) = \mathbb{C} \setminus \sigma(\mathcal{H})$, where $\rho(\mathcal{H})$ denotes the resolvent set of $\mathcal{H}$ and $\mathbb{C}$ is the set of all complex numbers. $\mathcal{R}_{\mathcal{T}}(z)$ is holomorphic $\forall z \in \rho(\mathcal{H})$. For $z \in \rho(\mathcal{H})$, $\zeta \in H$ it follows

$$\mathcal{R}_{\mathcal{H}}(z) = \int_{\sigma(\mathcal{H})} \frac{1}{z-\lambda}\, d\mathcal{E}(\lambda)\,, \tag{C.1}$$

$$G_{\zeta,\zeta}(z) = (\mathcal{R}_{\mathcal{H}}(z)\eta, \eta) = \int_{\sigma(\mathcal{H})} \frac{1}{z-\lambda}\, d\mathcal{E}_{\zeta,\zeta}(\lambda). \tag{C.2}$$

Taking the series expansion of $G_{\zeta,\zeta}(z)$ around the point $z = \infty$, we obtain

$$\begin{aligned} G_{\zeta,\zeta}(z) &= \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \left(\mathcal{H}^k \eta, \eta\right) \\ &= \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \int_a^b \lambda^k \, d\mathcal{E}_{\zeta,\zeta}(\lambda) = \sum_{k=0}^{\infty} \frac{\mu_k}{z^{k+1}} \end{aligned} \tag{C.3}$$

which converges for $|z| > \|\mathcal{H}\|$.

Diagonal Padé approximants for $G_{\zeta,\zeta}(z)$ are rational functions of the form $q_K(z)/p_K(z)$ which satisfy the following property

$$\left(G_{\zeta,\zeta}(z) - \frac{q_K(z)}{p_K(z)}\right) = \mathcal{O}\left(\frac{1}{z^{2K+1}}\right) \tag{C.4}$$

and are locally the best rational approximations for a given power series like Eqn. (C.3) (cf. e.g. Suetin (2002)). The denominator polynomials $p_K(z)$, are a set of orthogonal polynomials with respect to the measure $\mathcal{E}_{\eta,\eta}$ (cf. e.g. Moren and Branquinho (2008))

$$\int_a^b \lambda^k p_K(\lambda) \, d\mathcal{E}_{\zeta,\zeta}(\lambda) = 0\,, \quad k = 0, 1, \ldots, K-1 \tag{C.5}$$

and the numerator polynomial $q_K(z)$ can be expressed in terms of $p_K(z)$ as follows

$$q_K(z) = \int_a^b \frac{p_K(z) - p_K(\lambda)}{z - \lambda} \, d\mathcal{E}_{\zeta,\zeta}(\lambda). \tag{C.6}$$

The Padé approximants are constructed using the coefficients of the power series and provide an efficient analytic continuation of the series beyond the domain of convergence, which in the

present case is given by $|z| > \|\mathcal{H}\|$. The convergence of diagonal Padé approximants follows from Markov's theorem (cf. e.g. Suetin (2002)), whereby any Markov function like $G_{\zeta,\zeta}(z)$ can be recovered in $\rho(\mathcal{H})$ from the coefficients of the Laurent expansion of the function at the point $z = \infty$ (i.e. from the moments of the measure $\mathcal{E}_{\zeta,\zeta}$).

The orthogonal polynomials $\{p_k\}_{k=1}^{K}$ can be generated via the recurrence relation (cf. e.g. Golub and Meurant (2010))

$$p_{k+1}(\lambda) = (\lambda - a_{k+1})p_k(\lambda) - b_k^2 p_{k-1}(\lambda)\,, \quad k = 0, 1, \ldots, K-1$$
$$p_{-1}(\lambda) = 0\,, \quad p_0(\lambda) = 1 \tag{C.7}$$

where

$$\begin{aligned} a_{k+1} &= \frac{\langle \lambda p_k, p_k \rangle_\zeta}{\langle p_k, p_k \rangle_\zeta}\,, \quad k = 0, 1, \ldots, K-1 \\ b_k^2 &= \frac{\langle p_k, p_k \rangle_\zeta}{\langle p_{k-1}, p_{k-1} \rangle_\zeta}\,, \quad k = 1, 2, \ldots, K-1. \end{aligned} \tag{C.8}$$

Using Eqns. C.5 and C.6, we see that $\{q_k\}_{k=1}^{K}$ can be generated with a similar recurrence relation as that for $\{p_k\}_{k=1}^{K}$ but with different initializing conditions

$$q_{k+1}(\lambda) = (\lambda - a_{k+1})q_k(\lambda) - b_k^2 q_{k-1}(\lambda)\,, \quad k = 0, 1, \ldots, K-1$$
$$q_{-1}(\lambda) = -1,\ q_0(\lambda) = 0 \text{ and } b_0^2 = 1. \tag{C.9}$$

Corresponding to these recurrence relations we have the tridiagonal matrix

$$J_K = \begin{pmatrix} a_1 & 1 & & & \\ b_1^2 & a_2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & b_{K-2}^2 & a_{K-1} & 1 \\ & & & b_{K-1}^2 & a_K \end{pmatrix}. \tag{C.10}$$

Expanding $\det(zI_k - J_k)$ with respect to its last row, we obtain the recurrence formula for the determinant

$$\det(zI_{k+1} - J_{k+1}) = (z - a_{k+1})\det(zI_k - J_k) - b_k^2 \det(zI_{k-1} - J_{k-1})\,, k = 0, 1, \ldots, K-1$$

$$\det(zI_{-1} - J_{-1}) = 0\,, \quad \det(zI_0 - J_0) = 1$$

which is the same as that satisfied by $\{p_k\}_{k=0}^{K-1}$ as well as $\{\det(zI_k - \hat{J}_k)\}_{k=0}^{K-1}$. Consequently

$$p_K(z) = \det(zI_K - J_K) = \det(zI_K - \hat{J}_K). \tag{C.11}$$

Similarly

$$q_K(z) = \det(zI_{K-1} - J_{K-1}^1) = \det(zI_{K-1} - \hat{J}_{K-1}^1) \tag{C.12}$$

where the superscript i is used to denote the matrix obtained by deleting the first i rows and

columns. Therefore

$$\frac{q_K(z)}{p_K(z)} = \frac{\det(zI_{K-1} - \hat{J}^1_{K-1})}{\det(zI_K - \hat{J}_K)} = e_1(zI_K - \hat{J}_K)^{-1}e_1. \tag{C.13}$$

Above, the first column of I_K is denoted by e_1.

Expanding $\det(zI_K - \hat{J}_K)$ about the first row or column we obtain from Eqn. (C.13)

$$\frac{q_K(z)}{p_K(z)} = \cfrac{1}{z - a_1 - \cfrac{\det(zI_K - 2 - \hat{J}^2_{K-2})}{\det(zI_K - 2 - \hat{J}^1_{K-2})}} \tag{C.14}$$

and continuing similarly we obtained a continued fraction representation for the Padé approximant

$$\frac{q_K(z)}{p_K(z)} = \cfrac{1}{z - a_1 - \cfrac{b_1^2}{z - a_2 - \ldots - \cfrac{b_{K-1}^2}{z - a_K}}} \tag{C.15}$$

In the recursion method, the above expression is used to find the local density of states (Haydock (1980)) from which all the necessary quantities are evaluated. Since the local density of states has poles at the zeros of $p_K(z)$, a number of techniques to smoothen it have been developed. These include either using a terminator, the simplest one being $\{a_k\}_{k=K+1}^{\infty} = a_\infty$, $\{b_k\}_{k=K}^{\infty} = b_\infty$ or evaluating the Padé approximant slightly away from the real axis. However, all these are uncontrolled approximations and can lead to serious inaccuracies.

Finally, we look at the connection between Padé approximation and Gauss quadrature (cf. e.g. Van Assche (2006)). Multiplying both sides of Eqn. (C.4) by a polynomial of degree at most $2K-1$ denoted by $\pi_{2K-1}(z)$, integrating along a contour Γ encircling the real line and subsequently using Eqn. (C.2) we obtain

$$\int_a^b \pi_{2K-1}(\lambda)\, d\mathcal{E}_{\eta,\eta}(\lambda) = \sum_{j=1}^{K} w_j \pi_{2K-1}(\lambda_j) \tag{C.16}$$

where

$$w_j = \frac{q_K(\lambda_j)}{p_K^{'}(\lambda_j)} \tag{C.17}$$

is the residue of the Padé approximant at the zeros λ_j of p_K. It is clear that all polynomials of degree $2K-1$ are integrated exactly, therefore establishing the connection between the Padé approximation, recursion method and the LSSGQ method.

Appendix D

LSSGQ method with the finite-difference approximation

On using the finite-difference approximation for the Laplacian, we end up with the following matrix eigenvalue problem

$$\mathbf{H}\boldsymbol{\psi}_n = \lambda_n \boldsymbol{\psi}_n \,, \quad n = 1, 2, \ldots, N_d \tag{D.1}$$

where $\boldsymbol{\psi}_n = [\psi_{n,1}, \psi_{n,2}, \ldots, \psi_{n,N_d}]^T$ is a vector of wavefunction values at the uniformly spaced finite-difference nodes $\{\mathbf{x}_p\}_{p=1}^{N_d}$. The Lanczos iteration (Eqn. (4.48)) takes the form

$$\begin{aligned} b_{k+1}\mathbf{v}_{k+1} &= (\mathbf{H} - a_{k+1})\mathbf{v}_k - b_k \mathbf{v}_{k-1} \,, \;\; k = 0, 1, \ldots, K-1 \\ \mathbf{v}_{-1} &= \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix} \,, \quad \mathbf{v}_0 = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \,, \quad b_0 = 1 \end{aligned} \tag{D.2}$$

where

$$a_{k+1} = \mathbf{v}_k \mathbf{H} \mathbf{v}_k \,, \quad k = 0, 1, \ldots, K-1 \tag{D.3}$$

and b_k is computed such that $\|\mathbf{v}_k\|_{l^2} = 1$. For the p^{th} finite-difference node, only the p^{th} component of the initial vector $\mathbf{v}_0$ is nonzero and is set to unity. The calculation of the spectral quadrature nodes $\{\lambda_k^p\}_{k=1}^K$ and weights $\{w_k^p\}_{k=1}^K$ for each node $p \in [1, N_d]$ proceeds as described in Section 4.3.2. Subsequently, we can evaluate the Fermi energy by solving for the constraint

$$N_e = h^D \sum_{p=1}^{N_d} \rho^p \tag{D.4}$$

where

$$\rho^p = \frac{2}{h^D} \sum_{k=1}^{K} w_k^p g(\lambda_k^p, \lambda_f) \tag{D.5}$$

is the electron density at the p^{th} finite-difference node. Further, the band structure energy and entropy can be evaluated

$$U_{band} = \sum_{p=1}^{N_d} u^p \tag{D.6}$$

$$S = \sum_{p=1}^{N_d} s^p \tag{D.7}$$

where

$$u^p = 2\sum_{k=1}^{K} w_k^p \lambda_k^p g(\lambda_k^p, \lambda_f) \tag{D.8}$$

$$s^p = -2k_B \sum_{k=1}^{K} w_k^p g(\lambda_k^p, \lambda_f) \log g(\lambda_k^p, \lambda_f) + (1 - g(\lambda_k^p, \lambda_f) \log(1 - g(\lambda_k^p, \lambda_f))). \tag{D.9}$$

Above, h denotes the finite-difference nodal spacing and D is the spatial dimension.

Appendix E

LSSGQ method applied to systems with periodicity

In this section, we discuss the application of the LSSGQ method in the study of periodic systems. Typically, this would involve the need for Bloch periodic boundary conditions on the wavefunction. However, since we directly evaluate the electron density and not the individual wavefunctions, we can circumvent the need for Bloch periodic boundary conditions using the procedure described below. For clarity, we discuss the procedure in terms of the finite-difference approximation. However, the method is not restricted by the choice of basis.

First, we specify a representative volume Ω_{RV} which has the property that the solution for the entire system can be ascertained through a mapping of the solution obtained in Ω_{RV}. We use a finite-difference grid of uniform spacing h to discretize Ω_{RV}, which we denote by χ_{RV}. Next, we define an extended volume $\Omega_{EV} \supset \Omega_{RV}$, which is discretized using a finite-difference grid of spacing h, denoted by $\chi_{EV} \supset \chi_{RV}$. We evaluate the spectral quadrature nodes and weights for χ_{RV} using the procedure outlined in Appendix D. It is important to note that the size of Ω_{EV} is chosen such that all the vectors in the Lanczos iteration given by Eqn. (D.2) have nonzero components only for finite-difference nodes inside Ω_{EV}. Since the initial vector $\mathbf{v_0}$ always has a single nonzero

entry, the required size of Ω_{EV} can be easily calculated using the bandwidth of the discretized $\mathcal{H}$ and the number of spectral quadrature points.

Bibliography

Anantharaman, A., Cances, E., 2008. On Kohn-Sham models with LDA and GGA exchange-correlation functionals.
URL http://arxiv.org/abs/0809.5139v1

Arroyo, M., Ortiz, M., 2006. Local maximum-entropy approximation schemes: a seamless bridge between finite elements and meshfree methods. International Journal for Numerical Methods in Engineering 65, 2167–2202.

Bachelet, G. B., Hamann, D. R., Schlüter, M., 1982. Pseudopotentials that work: From H to Pu. Physical Review B 26 (8), 4199–4228.

Barrault, M., Cances, E., Hager, W., Bris, C. L., 2007. Multilevel domain decomposition for electronic structure calculations. Journal of Computational Physics 222 (1), 86 – 109.

Bernstein, N., Kermode, J. R., Csanyi, G., 2009. Hybrid atomistic simulation methods for materials systems. Reports on Progress in Physics 72 (2), 026501.

Bey, J., Aachen, R., 2000. Simplicial grid refinement: on Freudenthal's algorithm and the optimal number of congruence classes. Numerische Mathematik 85, 1–29.

Bompadre, A., Ortiz, M., 2010. Convergence analysis of local maximum-entropy approximation schemes. In preparation.

Bowler, D., Choudhury, R., Gillan, M., Miyazaki, T., 2006. Recent progress with large-scale ab initio calculations: the CONQUEST code. Physica Status Solidi B-Basic Solid State Physics 243 (5), 989–1000.

Brown, P. N., 1987. A local convergence theory for combined inexact-newton/finite-difference projection methods. SIAM Journal on Numerical Analysis 24 (2), 407–434.

Brown, P. N., Saad, Y., 1990. Hybrid Krylov methods for nonlinear systems of equations. SIAM Journal on Scientific and Statistical Computing 11 (3), 450–481.

Cances, E., Bris, C. L., Lions, P.-L., 2008. Molecular simulation and related topics: some open mathematical problems. Nonlinearity 21 (9), T165.

Car, R., Parrinello, M., Nov 1985. Unified approach for molecular dynamics and density-functional theory. Phys. Rev. Lett. 55 (22), 2471–2474.

Castro, A., Appel, H., Oliveira, M., Rozzi, C. A., Andrade, X., Lorenzen, F., Marques, M. A. L., Gross, E. K. U., Rubio, A., 2006. octopus: a tool for the application of time-dependent density functional theory. Physica Status Solidi B-Basic Solid State Physics 243 (11), 2465–2488.

Ceperley, D. M., Alder, B. J., 1980. Ground state of the electron gas by a stochastic method. Physical Review Letters 45 (7), 566–569.

Chelikowsky, J. R., Troullier, N., Saad, Y., 1994. Finite-difference-pseudopotential method: Electronic structure calculations without a basis. Physical Review Letters 72 (8), 1240–1243.

Choly, N., Lu, G., E, W., Kaxiras, E., Mar 2005. Multiscale simulations in simple metals: A density-functional-based methodology. Phys. Rev. B 71 (9), 094101.

Ciarlet, P., 2002. The finite-element method for elliptic problems. SIAM, Philadelphia.

Cyron, C. J., Arroyo, M., Ortiz, M., SEP 24 2009. Smooth, second order, non-negative meshfree approximants selected by maximum entropy. International Journal for Numerical Methods in Engineering 79 (13), 1605–1632.

Dal Maso, G., 1993. An introduction to Γ-convergence. Birkhäuser, Boston.

Dennis, J., Schnabel, R., 1996. Numerical Methods for Unconstrained Optimization and Nonlinear Equations. SIAM, Philadelphia.

Engel, E., Höck, A., Schmid, R. N., Dreizler, R. M., Chetty, N., 2001. Role of the core-valence interaction for pseudopotential calculations with exact exchange. Physical Review B 64 (12), 125111.

Fang, H.-r., Saad, Y., MAR 2009. Two classes of multisecant methods for nonlinear acceleration. Numerical Linear Algebra with Applications 16 (3), 197–221.

Finnis, M., 2003. Interatomic forces in condensed matter. Oxford University Press.

Fiolhais, C., Nogueira, F., Henriques, C., 1996. Evanescent core pseudopotential: Applications to

surfaces and clusters. Progress in Surface Science 53 (2-4), 315 – 322, proceedings of the 18th International Seminar on Surface Physics.

Fiolhais, C., Perdew, J. P., Armster, S. Q., MacLaren, J. M., Brajczewska, M., 1995. Dominant density parameters and local pseudopotentials for simple metals. Physical Review B 51 (20), 14001–14011.

Forsythe, G., Malcolm, M., Moler, C., 1973. Computer methods for mathematical computations. Prentice-Hall.

Freund, R. W., 1993. A transpose-free quasi-minimal residual algorithm for non-hermitian linear systems. SIAM Journal on Scientific Computing 14 (2), 470–482.

Galli, G., Parrinello, M., Dec 1992. Large scale electronic structure calculations. Phys. Rev. Lett. 69 (24), 3547–3550.

García-Cervera, C. J., Lu, J., E, W., Dec 2007. A sub-linear scaling algorithm for computing the electronic structure of materials. Communications in Mathematical Sciences 5 (4), 999–1026.

Garcia-Cervera, C. J., Lu, J., Xuan, Y., E, W., 2009. Linear-scaling subspace-iteration algorithm with optimally localized nonorthogonal wave functions for Kohn-Sham density functional theory. Physical Review B (Condensed Matter and Materials Physics) 79 (11), 115110.

García-Cervera, C. J., Lu, J., Xuan, Y., E, W., Mar 2009. Linear-scaling subspace-iteration algorithm with optimally localized nonorthogonal wave functions for Kohn-Sham density functional theory. Phys. Rev. B 79 (11), 115110.

Gavini, V., Bhattacharya, K., Ortiz, M., 2007a. Quasi-continuum orbital-free density-functional theory: A route to multi-million atom non-periodic DFT calculation. Journal of the Mechanics and Physics of Solids 55 (4), 697 – 718.

Gavini, V., Knap, J., Bhattacharya, K., Ortiz, M., 2007b. Non-periodic finite-element formulation of orbital-free density functional theory. Journal of the Mechanics and Physics of Solids 55 (4), 669 – 696.

Gilbarg, D., Trudinger, N., 1983. Elliptic differential equations of second order. Springer, New York.

Gillan, M. J., 1989. Calculation of the vacancy formation energy in aluminium. Journal of Physics: Condensed Matter 1 (4), 689.

Goedecker, S., Jul 1999. Linear scaling electronic structure methods. Rev. Mod. Phys. 71 (4), 1085–1123.

Golub, G., Meurant, G., 2010. Matrices, Moments and Quadrature with Applications. Princeton University Press.

Gonze, X., Beuken, J. M., Caracas, R., Detraux, F., Fuchs, M., Rignanese, G. M., Sindic, L., Verstraete, M., Zerah, G., Jollet, F., Torrent, M., Roy, A., Mikami, M., Ghosez, P., Raty, J. Y., Allan, D. C., 2002. First-principles computation of material properties: the abinit software project. Computational Materials Science 25, 478–492(15).

Govind, N., Wang, Y. A., Carter, E. A., 1999. Electronic-structure calculations by first-principles

density-based embedding of explicitly correlated systems. The Journal of Chemical Physics 110 (16), 7677–7688.

Gygi, F. m. c., Galli, G., Jul 1995. Real-space adaptive-coordinate electronic-structure calculations. Physical Review B 52 (4), R2229–R2232.

Haydock, R., 1980. Solid State Physics. Vol. 35. Academic Press.

Hehre, W. J., Stewart, R. F., Pople, J. A., 1969. Self-consistent molecular-orbital methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. The Journal of Chemical Physics 51 (6), 2657–2664.

Hohenberg, P., Kohn, W., 1964. Inhomogeneous electron gas. Physical Review 136 (3B), B864–B871.

Huber, K., 1972. Constants of diatomic molecules. McGraw-Hill, New York.

Ismail-Beigi, S., Arias, T. A., 2000. New algebraic formulation of density functional calculation. Computer Physics Communications 128 (1-2), 1 – 45.

Kleinman, L., Bylander, D. M., 1982. Efficacious form for model pseudopotentials. Physical Review Letters 48 (20), 1425–1428.

Knap, J., Ortiz, M., SEP 2001. An analysis of the quasicontinuum method. Journal of the Mechanics and Physics of Solids 49 (9), 1899–1923.

Kohn, W., 1999. Nobel lecture: Electronic structure of matter—wave functions and density functionals. Reviews of Modern Physics 71 (5), 1253–1266.

Kohn, W., Sham, L. J., 1965. Self-consistent equations including exchange and correlation effects. Physical Review 140 (4A), A1133–A1138.

Kotochigova, S., Levine, Z. H., Shirley, E. L., Stiles, M. D., Clark, C. W., 1997. Local-density-functional calculations of the energy of atoms. Physical Review A 55 (1), 191–199.

Kresse, G., Furthmüller, J., 1996. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Physical Review B 54 (16), 11169–11186.

Langreth, D. C., Mehl, M. J., 1983. Beyond the local-density approximation in calculations of ground-state electronic properties. Physical Review B 28 (4), 1809–1834.

Lu, G., Tadmor, E. B., Kaxiras, E., 2006. From electrons to finite elements: A concurrent multi-scale approach for metals. Physical Review B (Condensed Matter and Materials Physics) 73 (2), 024108.

Martin, R., 2004. Electronic Structure: Basic theory and practical methods. Cambridge University Press.

Mauri, F., Galli, G., Car, R., Apr 1993. Orbital formulation for electronic-structure calculations with linear system-size scaling. Phys. Rev. B 47 (15), 9973–9976.

Modine, N. A., Zumbach, G., Kaxiras, E., Apr 1997. Adaptive-coordinate real-space electronic-structure calculations for atoms, molecules, and solids. Physical Review B 55 (16), 10289–10301.

Moren, A., Branquinho, A., 2008. Coimbra lecture notes on Orthogonal Polynomials. Nova Science Publishers.

Nogueira, F., Fiolhais, C., He, J., Perdew, J. P., Rubio, A., 1996. Transferability of a local pseudopotential based on solid-state electron density. Journal of Physics: Condensed Matter 8 (3), 287–302.

Parr, R., Yang, W., 1989. Density-functional theory of atoms and molecules. Oxford University Press.

Pask, J. E., Klein, B. M., Fong, C. Y., Sterne, P. A., 1999. Real-space local polynomial basis for solid-state electronic-structure calculations: A finite-element approach. Physical Review B 59 (19), 12352–12358.

Pask, J. E., Sterne, P. A., 2005. Finite element methods in ab initio electronic structure calculations. Modelling and Simulation in Materials Science and Engineering 13 (3), R71–R96.

Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A., Joannopoulos, J. D., 1992. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. Reviews of Modern Physics 64 (4), 1045–1097.

Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R., Singh, D. J., Fiolhais, C., 1992. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. Physical Review B 46 (11), 6671–6687.

Perdew, J. P., Wang, Y., 1992. Accurate and simple analytic representation of the electron-gas correlation energy. Physical Review B 45 (23), 13244–13249.

Pickett, W. E., 1989. Pseudopotential methods in condensed matter applications. Computer Physics reports 9 (3), 115 – 197.

Rappe, A. M., Rabe, K. M., Kaxiras, E., Joannopoulos, J. D., 1990. Optimized pseudopotentials. Physical Review B 41 (2), 1227–1230.

Roothaan, C. C. J., 1951. New developments in molecular orbital theory. Reviews of Modern Physics 23 (2), 69–89.

Rosolen, A., Millan, D., Arroyo, M., MAY 14 2010. On the optimum support size in meshfree methods: A variational adaptivity approach with maximum-entropy approximants. International Journal for Numerical Methods in Engineering 82 (7), 868–895.

Rudin, W., 1991. Functional Analysis. McGraw Hill.

Saad, Y., 1992. Numerical Methods for Large Eigenvalue Problems. Manchester University Press.

Saad, Y., Schultz, M. H., 1986. GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems. SIAM Journal on Scientific and Statistical Computing 7 (3), 856–869.

Segall, M. D., Lindan, P. J. D., Probert, M. J., Pickard, C. J., Hasnip, P. J., Clark, S. J., Payne, M. C., 2002. First-principles simulation: ideas, illustrations and the CASTEP code. Journal of Physics: Condensed Matter 14 (11), 2717–2744.

Skylaris, C.-K., Haynes, P. D., Mostofi, A. A., Payne, M. C., 2005. Introducing ONETEP: Linear-

scaling density functional simulations on parallel computers. The Journal of Chemical Physics 122 (8), 084119.

Soler, J. M., Artacho, E., Gale, J. D., García, A., Junquera, J., Ordejón, P., Sánchez-Portal, D., 2002. The siesta method for ab initio order-n materials simulation. Journal of Physics: Condensed Matter 14 (11), 2745–2779.

Solin, P., S. K., Dolezel, I., 2004. Higher-order Finite Element methods. Chapman and Hall.

Struwe, M., 1990. Variational Methods, Applications to Nonlinear Partial Differential Equations and Hamiltonian Systems. Springer, Berlin.

Suetin, S. P., 2002. Padé approximants and efficient analytic continuation of a power series. Russian Mathematical Surveys 57 (1), 43.

Suryanarayana, P., Bhattacharya, K., Ortiz, M., 2011. A mesh-free convex approximation scheme for kohn-sham density functional theory. Journal of Computational Physics 230 (13), 5226 – 5238.

Suryanarayana, P., Gavini, V., Blesgen, T., Bhattacharya, K., Ortiz, M., 2010. Non-periodic finite-element formulation of Kohn-Sham density functional theory. Journal of the Mechanics and Physics of Solids 58 (2), 256 – 280.

Tadmor, E., Ortiz, M., Phillips, R., JUN 1996. Quasicontinuum analysis of defects in solids. Philosophical Magazine A-Physics of Condensed Matter Structure Defects and Mechanical Properties 73 (6), 1529–1563.

Thoutireddy, P., 2002. Variational arbitrary Lagrangian-Eulerian method. Ph.D. thesis, California Institute of Technology.

Troullier, N., Martins, J. L., 1991. Efficient pseudopotentials for plane-wave calculations. Physical Review B 43 (3), 1993–2006.

Tsuchida, E., 2004. Ab initio molecular-dynamics study of liquid formamide. The Journal of Chemical Physics 121 (10), 4740–4746.

Van Assche, W., 2006. Padé and hermite-padé approximation and orthogonality. Surveys in Approximation Theory (2006) 2, 61–91.

van der Vorst, H. A., 1992. BI-CGSTAB: a fast and smoothly converging variant of BI-CG for the solution of nonsymmetric linear systems. SIAM Journal on Scientific and Statistical Computing 13 (2), 631–644.

Vanderbilt, D., 1990. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. Physical Review B 41 (11), 7892–7895.

Veillard, A., Clementi, E., 1968. Correlation energy in atomic systems. v. degeneracy effects for the second-row atoms. The Journal of Chemical Physics 49 (5), 2415–2421.

Vogel, C. R., 2002. Computational methods for inverse problems. Society for Industrial and Applied Mathematics.

Weinert, M., Davenport, J. W., Jun 1992. Fractional occupations and density-functional energies and forces. Phys. Rev. B 45 (23), 13709–13712.

Wills, J. M., Cooper, B. R., 1987. Synthesis of band and model hamiltonian theory for hybridizing cerium systems. Physical Review B 36 (7), 3809–3823.

Zumbach, G., Modine, N. A., Kaxiras, E., 1996. Adaptive coordinate, real-space electronic structure calculations on parallel computers. Solid State Communications 99 (2), 57 – 61.